

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 00:35:50 ; Search time 1937.48 Seconds
(without alignments)
5426.703 Million cell updates/sec

Title: US-09-809-524B-1
Perfect score: 779
Sequence: 1 atggcacacaggtataatga.....gactggtttccatctgtatg 779

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	47.4	6.1	788	12 CNS00AK8	AL055582 Drosophila
C 2	47.2	6.1	903	12 CNS00IIC	AL074609 Drosophila
C 3	46.2	5.9	240	9 AU071288	AU071288 AU071288
C 4	46.2	5.9	240	9 AU072338	AU072338 AU072338
C 5	46.2	5.9	653	10 C24613	C24613 C24613 Dict
C 6	45.2	5.8	878	12 CNS0187R	AL108993 Drosophila
C 7	44.4	5.7	568	12 CNS010PL	AL099267 Drosophila
C 8	43.8	5.6	1101	12 CNS00RK2	AL077673 Drosophila
C 9	43.6	5.6	1053	12 CNS016CV	AL106585 Drosophila
C 10	41.8	5.4	908	12 CNS04KS8	AL295217 Tetraodon
C 11	41.8	5.4	939	12 CNS00CNG	AL059400 Drosophila
C 12	41.8	5.4	1300	10 BM468018	BM468018 AGENCOURT
C 13	41.4	5.3	536	12 AZ385473	AZ385473 IM0144B03
C 14	41.2	5.3	888	12 CNS027EX	AL184578 Tetraodon
C 15	40.8	5.2	421	9 AV865711	AV865711 AV865711
C 16	40.8	5.2	453	9 AV865286	AV865286 AV865286
C 17	40.6	5.2	711	12 CNS011VB	AL100766 Drosophila

C 18	40.6	5.2	842	12 CNS04993	AL280272 Tetraodon
C 19	40.6	5.2	1100	12 CNS012T7	AL101989 Drosophila
C 20	40.4	5.2	511	12 TAIL6D020	AL463945 T. brucei
C 21	40.4	5.2	996	12 CNS00FOH	AL071063 Drosophila
C 22	40.2	5.2	439	12 CNS017HP	AL108055 Drosophila
C 23	39.8	5.1	694	12 BH502270	BH502270 BOGIG53TR
C 24	39.8	5.1	969	12 CNS00IDL	AL074848 Drosophila
C 25	39.6	5.1	264	9 AV888582	AV888582 AV888582
C 26	39.6	5.1	1101	12 CNS001FK	AL060842 Drosophila
C 27	39.4	5.1	604	9 BE224195	BE224195 kp57f08.y
C 28	39.4	5.1	1101	12 CNS00LOO	AL068607 Drosophila
C 29	39.4	5.1	1147	12 B13042	B13042 T30M24-Sp6.
C 30	39.2	5.0	668	9 AL514901	AL514901 AL514901
C 31	39.2	5.0	823	12 CNS010SK	AL098546 Drosophila
C 32	39.2	5.0	1101	12 CNS00B39	AL056368 Drosophila
C 33	39	5.0	517	10 BF193908	BF193908 245644 MA
C 34	39	5.0	518	10 BJ079195	BJ079195 BJ079195
C 35	39	5.0	563	12 AQ319368	AQ319368 RPC111-98
C 36	38.8	5.0	721	12 AZ197220	AZ197220 SP 1034_A
C 37	38.8	5.0	859	12 CNS004YY	AL055406 Drosophila
C 38	38.8	5.0	1007	12 CNS06X9S	AL419462 T3 end of
C 39	38.8	5.0	1084	12 CNS00Z5S	AL097258 Drosophila
C 40	38.8	5.0	1101	12 CNS00L6N	AL068136 Drosophila
C 41	38.6	5.0	463	9 AV595306	AV595306 AV595306
C 42	38.6	5.0	549	12 TAIL14H1Q	AL462988 T. brucei
C 43	38.6	5.0	769	12 CNS04PQW	AL301649 Tetraodon
C 44	38.6	5.0	838	12 CNS00960	AL053091 Drosophila
C 45	38.6	5.0	993	12 CNS005N6	AL060418 Drosophila

ALIGNMENTS

RESULT 1

CNS00AK8/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqroff@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammeter in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's

pl and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers

1. 788

source

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

CNS00AK8 788 bp DNA linear GSS 04-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR21G10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

AL055582

AL055582.1 GI:4936452

GSS.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 788)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqroff@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammeter in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's

pl and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.


```
Query Match      5.9%; Score 46.2; DB 9; Length 240;
Best Local Similarity 55.2%; Pred. No. 2.7;
Matches 90; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 5 cacacaggtaaatgagtggttattgcttgctggtgctgttttttaccgacattaaagcc 64
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 CACAATTAAACAAAGAGATCTTGTATGCAATGGATGCTGTAACCTGGAAACATCAAGAC 118

QY 65 ttcttttattactccctattaaaaagtgccagtttagggcgctctttttacgtaagc 124
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 TTGCTCTAATTCGTAATAATGATTAAAAAGTTAGTATCGGTACTGATATTTCAAATATTT 178

QY 125 gcgtgcttgctgacttacccttccctattaccacatcattta 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 CAATGCCTGGTCTTTTATCCTTGCCAAATCAACTCTAAGTTA 221

RESULT 4
AU072338      240 bp mRNA linear EST 23-JUN-1999
LOCUS AU072338 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
DEFINITION dictyostelium cdna clone SSE223, mRNA sequence.
ACCESSION AU072338
VERSION AU072338.1 GI:5162527
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 240)
AUTHORS Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
TITLES Urushihara,H.
JOURNAL Developmental cDNA in Dictyostelium discoideum (1999)
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@sakura.cc.tsukuba.ac.jp
PROJECT = Dictyostelium discoideum cDNA project in Japan.

FEATURES
source
location/Qualifiers
1..240
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSE223"
/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slug"

BASE COUNT 90 a 36 c 42 g 72 t
ORIGIN

Query Match      5.9%; Score 46.2; DB 9; Length 240;
Best Local Similarity 55.2%; Pred. No. 2.7;
Matches 90; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 5 cacacaggtaaatgagtggttattgcttgctggtgctgtttttaccgacattaaagcc 64
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 CACAATTAAACAAAGAGATCTTGTATGCAATGGATGCTGTAACCTGGAAACATCAAGAC 125

QY 65 ttcttttattactccctattaaaaagtgccagtttagggcgctctttttacgtaagc 124
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 TTGCTCTAATTCGTAATAATGATTAAAAAGTTAGTATCGGTACTGATATTTCAAATATTT 185

QY 125 gcgtgcttgctgacttacccttccctattaccacatcattta 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 186 CAATGCCTGGTCTTTTATCCTTGCCAAATCAACTCTAAGTTA 228

RESULT 5
C24613      653 bp mRNA linear EST 28-APR-1999
LOCUS C24613 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
DEFINITION dictyostelium cdna clone SL-X008, mRNA sequence.
```

```
ACCESSION C24613
VERSION C24613.1 GI:2243034
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 653)
AUTHORS Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development
DNA Res. 5 (6), 335-340 (1998)
JOURNAL 99156227
MEDLINE
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@sakura.cc.tsukuba.ac.jp
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
POLYA-No. Location/Qualifiers
1..653
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SL-X008"
/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slug"

BASE COUNT 262 a 83 c 105 g 203 t
ORIGIN

Query Match      5.9%; Score 46.2; DB 10; Length 653;
Best Local Similarity 55.2%; Pred. No. 2.7;
Matches 90; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 5 cacacaggtaaatgagtggttattgcttgctggtgctgtttttaccgacattaaagcc 64
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 CACAATTAAACAAAGAGATCTTGTATGCAATGGATGCTGTAACCTGGAAACATCAAGAC 186

QY 65 ttcttttattactccctattaaaaagtgccagtttagggcgctctttttacgtaagc 124
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 TTGCTCTAATTCGTAATAATGATTAAAAAGTTAGTATCGGTACTGATATTTCAAATATTT 246

QY 125 gcgtgcttgctgacttacccttccctattaccacatcattta 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 CAATGCCTGGTCTTTTATCCTTGCCAAATCAACTCTAAGTTA 289

RESULT 6
CNS0187R/c
LOCUS CNS0187R
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN04E04 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL108993
VERSION AL108993.1 GI:5629297
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 878)
AUTHORS Drosophila melanogaster
TITLES Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
```


http://www.genoscope.cns.fr/Tetraodon.

```
FEATURES
source
1..908
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="117G15"
/clone_lib="G"
/note="Genoscope sequence ID : COBG117AD08SP1-end :
PUC-ori"
BASE COUNT 425 a 89 c 124 g 206 t 64 others
ORIGIN

Query Match 5.4%; Score 41.8; DB 12; Length 908;
Best Local Similarity 36.3%; Pred. No. 26;
Matches 160; Conservative 42; Mismatches 239; Indels 0; Gaps 0;

QY 338 tattcaattctacaatagaagctgaacacctcacttttggctgtctttcagccagtttt 397
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 591 TATTATTATTAATATTTTACTTCCAAATATATTTTWTWTWTWTWTWTWTWTWTWT 532
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 398 tgtgtgtatttcttataagcggcgaggtttatataacacattctgtatgagt 457
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 TTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTTC 472
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 458 catatcaattattaccaccaggcgctactttatttattgaccggcaatttttaaaatata 517
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 471 CTTTWTWTWTATWTWTGTTTCATATTTTATTTATTTATTTATTTATTTATTTATTT 412
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 518 tccaggcagagtggagaacgcttatacaattatgtcagtttctctctcctgcataa 577
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 TAATGPAWATWWTWCCTCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 352
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 578 tatgtatggtattagcagctcgtcttaggtcttttaaatcggctggcgcaacaattga 637
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 351 TWTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 292
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 638 atgtgttttcttcaatgcgcgtcaaaagtatatgttggttctactgaagctcctgatct 697
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291 WKTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 232
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 698 cattcccttatgctcttcatcactatttgggtgaagcgataaaattttatattatctaa 757
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 231 TTTAGCAAYAKACAGTTCACCKYKTKTKTAGYKGWKGKAYGTTTTTCCITTATTWACWT 172
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 758 aagaactggtttccatctgtat 778
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 TYCAVTGTTTTCCTTTTATTT 151
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
CNS00CNG/C 939 bp DNA linear GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR26H16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL059400
VERSION AL059400.1 GI:4946964
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 939)
AUTHORS Direct Submission
TITLE Determination of this BAC-end sequence was carried out as part of a
JOURNAL collaboration with the Berkeley Drosophila Genome Project (BDGP).
COMMENT - Web : www.genoscope.cns.fr
- The BDGP is constructing a physical map of the Drosophila
```

melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pleter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the HACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES
source
1..939
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR26H16"
/note="end : TET3"
BASE COUNT 71 a 349 c 104 g 180 t 235 others
ORIGIN

Query Match 5.4%; Score 41.8; DB 12; Length 939;
Best Local Similarity 9.7%; Pred. No. 26;
Matches 22; Conservative 124; Mismatches 81; Indels 0; Gaps 0;

QY 214 gggttagtcaccgagagtgattggttttttaattggtgtttgtcgcggttccc 273
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 585 GAGAGAGACGCGAKAKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 526
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 274 ttttggcggtgatagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 333
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 525 AKKAKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 466
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 334 acgatattcaattctacaatagaagctgaacctcacttttggctgtgtgtgtgtgtgt 393
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465 AKKAKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 406
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 394 ttttgtgtgtatttttttataagcgcggtgcatgagtttatatt 440
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 405 KDKAKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 359
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
BM468018/c 1300 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6437618 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5532913
DEFINITION 5', mRNA sequence.
ACCESSION BM468018
VERSION BM468018.1 GI:18517060
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1300)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12216 row: p column: 02
High quality sequence stop: 434.
Location/Qualifiers
1..1300
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 01:43:55 ; Search time 3427.07 Seconds
(without alignments)
4917.783 Million cell updates/sec

Title: US-09-809-524B-1
Perfect score: 779
Sequence: 1 atggcacacacaggttaaatga.....gactggttccatctgtatg 779

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 21979536 seqs, 10817449327 residues

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

- 1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
- 2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
- 3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
- 4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
- 5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
- 6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
- 7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
- 8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
- 9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
- 10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
- 11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
- 12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
- 13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
- 14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
- 15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
- 16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*
- 17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*
- 18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
- 19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
- 20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
- 21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
- 22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
- 23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
- 24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
- 25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
- 26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
- 27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
- 28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
- 29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
- 30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
- 31: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
- 32: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
- 33: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
- 34: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
- 35: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
- 36: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
- 37: /cgn2_6/ptodata/2/pna/US100_COMB.seq:*
- 38: /cgn2_6/ptodata/2/pna/US101_COMB.seq:*
- 39: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
- 40: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
- 41: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
- 42: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
- 43: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*

44: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*

45: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*

46: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*

47: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*

48: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*

49: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*

50: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*

51: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*

52: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*

53: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*

54: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*

55: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*

56: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*

57: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*

58: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*

59: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*

60: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*

61: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*

62: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*

63: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*

64: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*

65: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*

66: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*

67: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*

68: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*

69: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*

70: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*

71: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*

72: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*

73: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*

74: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*

75: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	779	100.0	779	31	US-09-809-524-1	Sequence 1, Appl
2	779	100.0	779	31	US-09-809-524A-1	Sequence 1, Appl
3	779	100.0	779	31	US-09-809-524B-1	Sequence 1, Appl
4	763	97.9	779	31	US-09-809-524-2	Sequence 2, Appl
5	763	97.9	779	31	US-09-809-524A-2	Sequence 2, Appl
6	763	97.9	779	31	US-09-809-524B-2	Sequence 2, Appl
7	750.6	96.4	13417	10	US-08-637-750-37	Sequence 37, Appl
8	750.6	96.4	13417	10	US-08-637-759A-37	Sequence 37, Appl
9	750.6	96.4	13417	12	US-08-871-355-37	Sequence 37, Appl
10	207.4	26.6	297	10	US-08-637-750-15	Sequence 15, Appl
11	207.4	26.6	297	10	US-08-637-759A-15	Sequence 15, Appl
12	207.4	26.6	297	12	US-08-871-355-15	Sequence 15, Appl
13	118	15.1	43360	18	US-09-453-702B-206	Sequence 206, App
14	118	15.1	45325	18	US-09-453-702B-261	Sequence 261, App
15	76	9.8	70559	18	US-09-409-800B-1	Sequence 1, Appl
16	63.6	8.2	810	21	US-09-543-681A-2303	Sequence 2303, Ap
17	60.2	7.7	801	33	US-09-868-604-33	Sequence 33, Appl
18	59.4	7.6	1086	16	US-09-252-991A-13855	Sequence 13855, A
19	59.4	7.6	1365	16	US-09-252-991A-13593	Sequence 13593, A
20	59.4	7.6	1740	16	US-09-252-991A-13753	Sequence 13753, A
21	59.2	7.6	24153	47	US-60-082-302-682	Sequence 682, App
22	53.6	6.9	16950	18	US-09-453-702B-166	Sequence 166, App
23	53	6.8	771	36	US-09-974-300-1867	Sequence 1867, Ap
24	52.2	6.7	184	10	US-08-637-759-16	Sequence 16, Appl
25	52.2	6.7	184	10	US-08-637-759A-16	Sequence 16, Appl
26	52.2	6.7	184	12	US-08-871-355-16	Sequence 16, Appl
27	48	6.2	2000	33	US-09-887-272A-5263	Sequence 5263, Ap
28	47.2	6.1	7218	8	US-08-466-194-14	Sequence 14, Appl
29	43.4	5.6	603	17	US-09-397-761A-2950	Sequence 2950, Ap
30	43.4	5.6	603	32	US-09-838-601-2950	Sequence 2950, Ap
31	41.4	5.3	625	27	US-09-680-598A-688	Sequence 688, App


```
Db 361 gaaacctcaacttttggcttcttccagccagtttttgggtgtattttctttataagc 420
QY 421 ggcggcatggagtttattataaacattctgtatgagtcataatattaccaccagg 480
Db 421 ggcggcatggagtttattataaacattctgtatgagtcataatattaccaccagg 480
QY 481 cgtactttattattgaccggcaatttttaaaatatatccaggcagagtgaggaaacgctt 540
Db 481 cgtactttattattgaccggcaatttttaaaatatatccaggcagagtgaggaaacgctt 540
QY 541 tatcaattatgtgtctctctctctccgcccaataatgtatggtattagccgactcg 600
Db 541 tatcaattatgtgtctctctctctccgcccaataatgtatggtattagccgactcg 600
QY 601 gcttaggtctcttttaaatcggtcgccacaaattggaatgtgtttcttctcaatgccg 660
Db 601 gcttaggtctcttttaaatcggtcgccacaaattggaatgtgtttcttctcaatgccg 660
QY 661 ctcaaaagtatatgttctactgacgctcctgatctcattcccttatgtctcttcacac 720
Db 661 ctcaaaagtatatgttctactgacgctcctgatctcattcccttatgtctcttcacac 720
QY 721 tatttgggtgaagcgataaatttttatatttataaaagactggtttcccatctgtatg 779
Db 721 tatttgggtgaagcgataaatttttatatttataaaagactggtttcccatctgtatg 779

RESULT 3
US-09-809-524B-1
; Sequence 1, Application US/09809524B
; GENERAL INFORMATION:
; APPLICANT: Lowery, David
; APPLICANT: Kennedy, Michael J
; TITLE OF INVENTION: Salmonella Vaccine Materials and Methods
; FILE REFERENCE: 28341/6114.N
; CURRENT APPLICATION NUMBER: US/09/809,524B
; CURRENT FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/190,178
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 779
; TYPE: DNA
; ORGANISM: Salmonella dublin
US-09-809-524B-1
```

```
Query Match 100.0%; Score 779; DB 31; Length 779;
Best Local Similarity 100.0%; Pred. No. 3.3e-186;
Matches 779; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 atggcacacacaggttaaatgagtggtcttatgcatggctgtggtctttattcgaccatta 60
Db 1 atggcacacacaggttaaatgagtggtcttatgcatggctgtggtctttattcgaccatta 60
QY 61 agcctttctttattacttccctattataaaagtggcagtttagggcgccgtctttacgt 120
Db 61 agcctttctttattacttccctattataaaagtggcagtttagggcgccgtctttacgt 120
QY 121 aatggcgctgcttatgtcaccttaccatattaccatattaccacagagaagatt 180
Db 121 aatggcgctgcttatgtcaccttaccatattaccatattaccacagagaagatt 180
QY 181 atgatgcataattgggtaaaagattacagttggttagggttagtcacccggagaggtattatt 240
Db 181 atgatgcataattgggtaaaagattacagttggttagggttagtcacccggagaggtattatt 240
QY 241 ggttttttaattgggttttgcgcggcggttcccttttgggcccgttgatatggcggggttt 300
Db 241 ggttttttaattgggttttgcgcggcggttcccttttgggcccgttgatatggcggggttt 300
```

```
QY 301 ctgcttgatactttacgtggcgacaaatgggtacgatatattcaattctcaataagaagct 360
Db 301 ctgcttgatactttacgtggcgacaaatgggtacgatatattcaattctcaataagaagct 360
QY 361 gaaacctcaacttttggcttcttccagccagtttttgggtgtattttctttataagc 420
Db 361 gaaacctcaacttttggcttcttccagccagtttttgggtgtattttctttataagc 420
QY 421 ggcggcatggagtttattataaacattctgtatgagtcataatattaccaccagg 480
Db 421 ggcggcatggagtttattataaacattctgtatgagtcataatattaccaccagg 480
QY 481 cgtactttattattgaccggcaatttttaaaatatatccaggcagagtgaggaaacgctt 540
Db 481 cgtactttattattgaccggcaatttttaaaatatatccaggcagagtgaggaaacgctt 540
QY 541 tatcaattatgtgtctctctctctccgcccaataatgtatggtattagccgactcg 600
Db 541 tatcaattatgtgtctctctctctccgcccaataatgtatggtattagccgactcg 600
QY 601 gcttaggtctcttttaaatcggtcgccacaaattggaatgtgtttcttctcaatgccg 660
Db 601 gcttaggtctcttttaaatcggtcgccacaaattggaatgtgtttcttctcaatgccg 660
QY 661 ctcaaaagtatatgttctactgacgctcctgatctcattcccttatgtctcttcacac 720
Db 661 ctcaaaagtatatgttctactgacgctcctgatctcattcccttatgtctcttcacac 720
QY 721 tatttgggtgaagcgataaatttttatatttataaaagactggtttcccatctgtatg 779
Db 721 tatttgggtgaagcgataaatttttatatttataaaagactggtttcccatctgtatg 779

RESULT 4
US-09-809-524-2
; Sequence 2, Application US/09809524
; GENERAL INFORMATION:
; APPLICANT: Pharmacia & Upjohn Company
; TITLE OF INVENTION: Salmonella Vaccine Materials and Methods
; FILE REFERENCE: 28341/6114.N
; CURRENT APPLICATION NUMBER: US/09/809,524
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/190,178
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 779
; TYPE: DNA
; ORGANISM: Salmonella typhimurium
US-09-809-524-2
```

```
Query Match 97.9%; Score 763; DB 31; Length 779;
Best Local Similarity 98.7%; Pred. No. 3.7e-182;
Matches 769; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
QY 1 atggcacacacaggttaaatgagtggtcttatgcatggctgtggtctttattcgaccatta 60
Db 1 atggcacacacaggttaaatgagtggtcttatgcatggctgtggtctttattcgaccatta 60
QY 61 agcctttctttattacttccctattataaaagtggcagtttagggcgccgtctttacgt 120
Db 61 agcctttctttattacttccctattataaaagtggcagtttagggcgccgtctttacgt 120
QY 121 aatggcgctgcttatgtcaccttaccatattaccatattaccacagagaagatt 180
Db 121 aatggcgctgcttatgtcaccttaccatattaccatattaccacagagaagatt 180
QY 181 atgatgcataattgggtaaaagattacagttggttagggttagtcacccggagaggtattatt 240
Db 181 atgatgcataattgggtaaaagattacagttggttagggttagtcacccggagaggtattatt 240
```

```
QY 241 ggtttttaaattggttttgcggcggttcccttttggcgcggttgatgagcggttt 300
|||||
Db 241 ggttttcaaattggttttgcggcggttcccttttggcgcggttgatgagcggttt 300
|||||
QY 301 ctgcttgatactttacgtggcgcgacaattgggtacgataattcaattctacatagaagct 360
|||||
Db 301 ctgcttgatactttacgtggcgcgacaattgggtacgataattcaattctacatagaagct 360
|||||
QY 361 gaaacctcaacttttggcttgcctttcagccagtttttgggtggttttctttataaagc 420
|||||
Db 361 gaaacctcaacttttggcttgcctttcagccagtttttgggtggttttctttataaagc 420
|||||
QY 421 ggcggcatgaggtttatataaacattctgtatgagtcataataatttaccaccagg 480
|||||
Db 421 ggcggcatgaggtttatataaacattctgtatgagtcataataatttaccaccagg 480
|||||
QY 481 cgtactttattattgacggcgaatttttaaaatatatccaggcagagtgagaaagcctt 540
|||||
Db 481 cgtactttattattgacggcgaatttttaaaatatatccaggcagagtgagaaagcctt 540
|||||
QY 541 tatcaattatgctcagtttctctctctcctgcataataatgtatgtattagccgactg 600
|||||
Db 541 tatcaattatgctcagtttctctctctcctgcataataatgtatgtattagccgactg 600
|||||
QY 601 gctttaggtcttttaaatgcgtgcgcgacaacaattggaatgtgttttcttctcaatgcg 660
|||||
Db 601 gctttaggtcttttaaatgcgtgcgcgacaacaattggaatgtgttttcttctcaatgcg 660
|||||
QY 661 ctcaaaagtatatattggttctactgacgctcctgcataatcccttatgctcttcacac 720
|||||
Db 661 ctcaaaagtatatattggttctactgacgctcctgcataatcccttatgctcttcacac 720
|||||
QY 721 tatttgggtgaaagcgataaattttatatttctaaagactggtttccatctgtatg 779
|||||
Db 721 tatttgggtgaaagcgataaattttatatttctaaagactggtttccatctgtatg 779
|||||

RESULT 5
US-09-809-524A-2
; Sequence 2, Application US/09809524A
; GENERAL INFORMATION:
; APPLICANT: Lowery, David
; APPLICANT: Kennedy, Michael J
; TITLE OF INVENTION: Salmonella Vaccine Materials and Methods
; FILE REFERENCE: 28341/6114.N
; CURRENT APPLICATION NUMBER: US/09/809,524A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/190,178
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 779
; TYPE: DNA
; ORGANISM: Salmonella typhimurium
US-09-809-524A-2

Query Match 97.9%; Score 763; DB 31; Length 779;
Best Local Similarity 98.7%; Pred. No. 3.7e-182;
Matches 769; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 atggcacacaggtataatgagtggttattgcttggtggttatttcgaccattta 60
|||||
Db 1 atggcacacaggtataatgagtggttattgcttggtggttatttcgaccattta 60
|||||
QY 61 agcctttctttattactccctattataaaagtggcagtttagggcgctctttta 120
|||||
Db 61 agcctttctttattactccctattataaaagtggcagtttagggcgctctttta 120
|||||
QY 121 aatggcgctgttatgtcacttacccttcccatattaccatattaccagcagaagatt 180
|||||
Db 121 aatggcgctgttatgtcacttacccttcccatattaccatattaccagcagaagatt 180
|||||
```

```
QY 181 atgatgcataattggttaaagattacagttggttaggttagtcaccggagaggtgattatt 240
|||||
Db 181 atgatgcataattggttaaagattacagttggttaggttagtcactggagaggtgattatt 240
|||||
QY 241 ggtttttaaattggttttgcggcggttcccttttggcgcggttgatgagcggtttt 300
|||||
Db 241 ggttttcaaattggttttgcggcggttcccttttggcgcggttgatgagcggtttt 300
|||||
QY 301 ctgcttgatactttacgtggcgcgacaattgggtacgataattcaattctacatagaagct 360
|||||
Db 301 ctgcttgatactttacgtggcgcgacaattgggtacgataattcaattctacatagaagct 360
|||||
QY 361 gaaacctcaacttttggcttgcctttcagccagtttttgggtggttttctttataaagc 420
|||||
Db 361 gaaacctcaacttttggcttgcctttcagccagtttttgggtggttttctttataaagc 420
|||||
QY 421 ggcggcatgaggtttatataaacattctgtatgagtcataataatttaccaccagg 480
|||||
Db 421 ggcggcatgaggtttatataaacattctgtatgagtcataataatttaccaccagg 480
|||||
QY 481 cgtactttattattgacggcgaatttttaaaatatatccaggcagagtgagaaagcctt 540
|||||
Db 481 cgtactttattattgacggcgaatttttaaaatatatccaggcagagtgagaaagcctt 540
|||||
QY 541 tatcaattatgctcagtttctctctcctgcataataatgtatgtattagccgactg 600
|||||
Db 541 tatcaattatgctcagtttctctctcctgcataataatgtatgtattagccgactg 600
|||||
QY 601 gctttaggtcttttaaatgcgtgcgcgacaacaattggaatgtgttttcttctcaatgcg 660
|||||
Db 601 gctttaggtcttttaaatgcgtgcgcgacaacaattggaatgtgttttcttctcaatgcg 660
|||||
QY 661 ctcaaaagtatatattggttctactgacgctcctgcataatcccttatgctcttcacac 720
|||||
Db 661 ctcaaaagtatatattggttctactgacgctcctgcataatcccttatgctcttcacac 720
|||||
QY 721 tatttgggtgaaagcgataaattttatatttctaaagactggtttccatctgtatg 779
|||||
Db 721 tatttgggtgaaagcgataaattttatatttctaaagactggtttccatctgtatg 779
|||||
```

```
RESULT 6
US-09-809-524B-2
; Sequence 2, Application US/09809524B
; GENERAL INFORMATION:
; APPLICANT: Lowery, David
; APPLICANT: Kennedy, Michael J
; TITLE OF INVENTION: Salmonella Vaccine Materials and Methods
; FILE REFERENCE: 28341/6114.N
; CURRENT APPLICATION NUMBER: US/09/809,524B
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/190,178
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 779
; TYPE: DNA
; ORGANISM: Salmonella typhimurium
US-09-809-524B-2

Query Match 97.9%; Score 763; DB 31; Length 779;
Best Local Similarity 98.7%; Pred. No. 3.7e-182;
Matches 769; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 atggcacacaggtataatgagtggttattgcttggtggttatttcgaccattta 60
|||||
Db 1 atggcacacaggtataatgagtggttattgcttggtggttatttcgaccattta 60
|||||
QY 61 agcctttctttattactccctattataaaagtggcagtttagggcgctctttta 120
|||||
Db 61 agcctttctttattactccctattataaaagtggcagtttagggcgctctttta 120
|||||
QY 121 aatggcgctgttatgtcacttacccttcccatattaccatattaccagcagaagatt 180
|||||
Db 121 aatggcgctgttatgtcacttacccttcccatattaccatattaccagcagaagatt 180
|||||
```

```
Db 61 agcctttttattacttcccttattataaaagtgccagtttagggcgccacttttacgt 120
QY 121 aatggcggtgcttatgtcaacttcccttccctattaccataattaccagcagaagatt 180
Db 121 aatggcggtgcttatgtcaacttcccttccctattaccataattaccagcagaagatt 180
QY 181 atgatgcataatgggtaaaagattacagttggttagggtagtcacccggagaggtgattatt 240
Db 181 atgatgcataatgggtaaaagattacagttggttagggtagtcacccggagaggtgattatt 240
QY 241 ggttttttaattgggtttttgtgcggcggttcccttttgggcgcttgatagggcggttt 300
Db 241 ggttttttaattgggtttttgtgcggcggttcccttttgggcgcttgatagggcggttt 300
QY 301 ctgcttgatactttacgtggcgacaaatgggtacgataattcaattctcaataagaagct 360
Db 301 ctgcttgatactttacgtggcgacaaatgggtacgataattcaattctcaataagaagct 360
QY 361 gaaacctcacttttggcttggcttttcagccagtttttggctgtgtattttttataagc 420
Db 361 gaaacctcacttttggcttggcttttcagccagtttttggctgtgtattttttataagc 420
QY 421 ggcggcgatggagttattataaaacattctgtatgatacatcaattattaccaccagg 480
Db 421 ggcggcgatggagttattataaaacattctgtatgatacatcaattattaccaccagg 480
QY 481 cgtactttattttgacccggcaaattttttaaaatatatccaggcagagtgagaaagcgtt 540
Db 481 cgtactttattttgacccggcaaattttttaaaatatatccaggcagagtgagaaagcgtt 540
QY 541 tatacaattatgtcagttctctctctctccgcaataatgatgtatgtattagccgatctg 600
Db 541 tatacaattatgtcagttctctctctctccgcaataatgatgtatgtattagccgatctg 600
QY 601 gcttaggtcttttaaatcggtgcgcacaaatggaatgtgttttctctcaatgcgg' 660
Db 601 gcttaggtcttttaaatcggtgcgcacaaatggaatgtgttttctctcaatgcgg' 660
QY 661 ctcaaaagataattggttctactgacgctctctgactctcattcccttatgctcttcacac 720
Db 661 ctcaaaagataattggttctactgacgctctctgactctcattcccttatgctcttcacac 720
QY 721 tatttgggtgaaagcgataaaattttattattattatctcaaaagactggttccatctgtatg 779
Db 721 tatttgggtgaaagcgataaaattttattattattatctcaaaagactggttccatctgtatg 779
```

RESULT 7

```
US-08-637-759-37
; Sequence 37, Application US/08637759
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center, 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,759
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
```

```
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: rpms 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13417 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: DNA sequence of VGC II from centre to left
; ORGANISM: hand end
; US-08-637-759-37
```

```
Query Match 96.4%; Score 750.6; DB 10; Length 13417;
Best Local Similarity 98.5%; Pred. No. 1.2e-178;
Matches 767; Conservative 1; Mismatches 10; Indels 1; Gaps 1;

QY 1 atggcacacacaggtaaatgagtgcttattgcatgtgctgtgctttatttcgaccatta 60
Db 9573 ATGGCACACACAGGTAAATGAGTGGCTTATTGCAATTGGCTGTGGCTTTATTTCGACCATTTG 9632
QY 61 agcctttctttattacttccccctattataaaagtgccagtttagggcgctctttcacgt 120
Db 9633 AGCCTTCTCTTTATTACTTCCCTTTATTAAGAGTGGCAGTTTAGGGCGCGCACATTTACGT 9692
QY 121 aatggcgctgttatgtcaacttaccctttcccatattaccataattaccatcatattaccagcagaagatt 180
Db 9693 AATGGCGTGTGTATGTCACTTACTTACCTTTCGGATATTACCAATCATTTACCAGCAGAGATT 9752
QY 181 atgatgcataatgggtaaaagattacagttggttagggtagtcacccggagaggtgattatt 240
Db 9753 ATGATGCATATTGGTAAAGATTACAGTTGGTTAGGTTAGTACACTGGAGAGGTGATTATT 9812
QY 241 ggttttttaattgggtttttgtgcggcggttcccttttgggcgcttgatagtcggcggtttt 300
Db 9813 GGTTTTTCAAATGGGTTTGTGGCGCGGTTCCTCTTTGGGCGCTTGATATCGCGGGGTTT 9872
QY 301 ctgcttgatactttacgtggcgacaaatgggtacgataattcaattctcaataagaagct 360
Db 9873 CTGCTTGATACTTTACGTGGCGCACAAATGGGTACGATATTCAATTCTACATAGAGCT 9932
QY 361 gaaacctcaactttttggcttggcttttcagccagtttttggctgtgtattttctttataagc 420
Db 9933 GAAACCTCACTTTTGGCTTGGCTTTTCAGCCAGTCTCTGTGTGTATTTCCTTTTATAAGC 9992
QY 421 ggcggcgatggagttattataaaacattctctgtatgagtcataatcaataattaccaccagg 480
Db 9993 GCGGCGCATGGAGTTTATATTAAACATTCGTATGAGTCATATCAATATTTACCACCAGGG 10052
QY 481 cgtactttattttgacccggcaaatttttataaaatatccaggcagagtgaggagaagcgtt 540
Db 10053 CGTACTTTTATTATTGACCAGCAATTTTAAATATATATCCAGGCAGAGTGGAGACGCTT 10112
QY 541 tatacaattatgtcagttctctctctctccctgcccataaataatgtatggttattagccgatctg 600
Db 10113 TATCAATTATGATACAGCTTCTCTCTCTGCCATAATATGATGTTATTATGCCGATCTG 10172
QY 601 gctttaggctcttttaaatcggtcggcacaaatggaatgtgtttttctctcctcaatgcgg 660
Db 10173 GCTTTAGGCTCTTTAAATCGGTCCGCGCACAAATTTGAATGTGTGTCTTCTTCAATGCCG 10232
QY 661 ctcaaaagataattggttctactgacgctccctgatctcattcccttatgctcttcacac 720
Db 10232 ctcaaaagataattggttctactgacgctccctgatctcattcccttatgctcttcacac 720
```

Db 10233 CTCAAAAGTATATTGGTTCTACTGAGG-YCCTGATCTCATTCCTTATGCTCTTCAAC 10291

QY 721 tatttggtgaaagcgaataattttattatttatactaaagacgggtttccatctgtatg 779
|||||

Db 10292 TATTGGTTGAAAGCGATAAATTTATATTTATCTAAAGACGTGGTTTCCATCTGTATG 10350

RESULT 8

US-08-637-759A-37

; Sequence 37, Application US/08637759A

; GENERAL INFORMATION:

; APPLICANT: David William Holden

; TITLE OF INVENTION: Identification of Genes

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Patrea L. Pabst

; STREET: 2800 One Atlantic Center

; CITY: Atlanta

; STATE: Georgia

; COUNTRY: USA

; ZIP: 30309-3450

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/637,759A

; FILING DATE: 03-MAY-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/GB95/02875

; FILING DATE: 11-DEC-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Pabst, Patrea L.

; REGISTRATION NUMBER: 31,284

; REFERENCE/DOCKET NUMBER: RPMS 101

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (404) 873-8794

; TELEFAX: (404) 873-8795

; INFORMATION FOR SEQ ID NO: 37:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 13417 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: DNA sequence of VGC II from centre to left

; ORGANISM: hand end

US-08-637-759A-37

Query Match 96.4%; Score 750.6; DB 10; Length 13417;

Best Local Similarity 98.5%; Pred. No. 1.2e-178;

Matches 767; Conservative 1; Mismatches 10; Indels 1; Gaps 1;

QY 1 atggcacaacaggttaaatgagtgcttattgcatggctgtggctttttatttcgacattta 60
|||||

Db 9573 ATGGCACACAGGTAAGTGGCTTATTGCTATTGCTGCTGGCTTTTATTCGACCATG 9632

QY 61 agcctttctttattacttccctattataaaagtgagtggttagggccgcctctttacgt 120
|||||

Db 9633 AGCCTTCTTATTACTTCCCTTATTAAAGTGGCAGTATTAGGGCCGCACTTTTACGT 9692

QY 121 aatggcgctgcttattgctacatttccctattataccaattattaccagagaagatt 180
|||||

Db 9693 AATGGCGTCTTATTGCTACTTACCTTTCCGATATTACCAATCATTTACCAGAGAAGATT 9752

QY 181 atgatcatattggttaagattacagtttggttaggttagtaccggagaggtgattatt 240
|||||

Db 9753 ATGATGCATATTGGTAAAGATTACAGTTGGTTAGGTTAGTCACTGGAGAGGTGATTATT 9812

QY 241 ggttttttaattggtttttgtgcgcggttccctttttggccggttgatatgcggggttt 300
|||||

Db 9813 GGTTTTCAATTGGTGTCTGTCGGCGGTTCCTTTTGGCGCGTTGATATGGCGGGGTTT 9872

QY 301 ctgcttgatacttttacgtggcgacaaatgggttacgatatcaattctacaatagaaagct 360
|||||

Db 9873 CTGCTTGATACTTTTACGTGGCGGACAAATGGGTACGATATTCAATTCTACATAGAAAGCT 9932

QY 361 gaaacctcaactttttggctgtcttttcagccagtttttgggtgtttattttcttttaagc 420
|||||

Db 9933 GAAACCTCACTTTTGGCTTGTCTTTCAGCCAGTCTTGTGTGTTATTCTTTTATAAGC 9992

QY 421 ggcggcatggagtttattataaacattctgtatgagtcataatcaattattaccacagg 480
|||||

Db 9993 GCGCGCATGGAGTTTATATAAACATTCTGTATGAGTCATATCAATATTACCACCAAGG 10052

QY 481 cgtactttattattgacggcgcaatttttaaaatatataatccaggcagagtgagagaacgctt 540
|||||

Db 10053 CGTACTTTATTTAGCCAGCAATTTTAAATATATATCCAGCAGAGTGGAGAACGCTT 10112

QY 541 tateaatattgctgaatttctctctccctgcataatcatgtattgattagccgacttg 600
|||||

Db 10113 TATCAATTATGATATCAGCTTCTCTCTCTCCGCCATAATATGATGTTATGATGATGATG 10172

QY 601 gctttagggttttaaatcggtcgggcacaacaattgaattggtttttcttctcaatgccg 660
|||||

Db 10173 GCTTTAGGTCTTTTAAATCGTGGCGCACAAATTTGAATGTGTTTCTTCTCAATGCCG 10232

QY 661 ctcaaaagtattggttctactgacgtcctctctctcattcccttatgctcttcacac 720
|||||

Db 10233 CTCAAAAGTATATTGGTTCTACTGAGG-YCCTGATCTCATTCCTTATGCTCTTCAAC 10291

QY 721 tatttggtgaaagcgaataattttattatttatactaaagacgtgggtttccatctgtatg 779
|||||

Db 10292 TATTGGTTGAAAGCGATAAATTTATATTTATCTAAAGACGTGGTTTCCATCTGTATG 10350

RESULT 9

US-08-871-355-37

; Sequence 37, Application US/08871355

; GENERAL INFORMATION:

; APPLICANT: David William Holden

; TITLE OF INVENTION: Identification of Genes

; NUMBER OF SEQUENCES: 38

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Patrea L. Pabst

; STREET: 2800 One Atlantic Center

; CITY: Atlanta

; STATE: Georgia

; COUNTRY: USA

; ZIP: 30309-3450

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/871,355

; FILING DATE: 09-JUN-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/GB95/02875

; FILING DATE: 11-DEC-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Pabst, Patrea L.

; REGISTRATION NUMBER: 31,284

; REFERENCE/DOCKET NUMBER: RPMS 101 CON

; TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 13417 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: DNA sequence of VGC II from centre to left
ORGANISM: hand end
US-08-871-355-37

Query Match 96.4%; Score 750.6; DB 12; Length 13417;
Best Local Similarity 98.5%; Pred. No. 1.2e-178;
Matches 767; Conservative 1; Mismatches 10; Indels 1; Gaps 1;

QY 1 atggcacacaggtataatgagtgcttattgcatggcttgctggtgctttattcgaccatt 60
Db 9573 ATGGCACACAGGTAATGAGTGGCTTATTGCAATGGCTGTGGCTTTATTGACCATTTG 9632
QY 61 agcctttcttattacttccctattataaaagtgccagtttagggcgctctttttacgt 120
Db 9633 AGCCTTCTTATTACTTCCCTTATTAAAGTGGCAGTTTAGGGCGGCACACTTTTACGT 9692
QY 121 aatggcggtgttatgtcacttacccttcccatattaccatatttaccagcagaagatt 180
Db 9693 AATGGCGTGTCTATGTCACCTTCCGATATTACCAATCATTTACCAGCAGAAGATT 9752
QY 181 atgatgcataattgtaaaagattacagattggttaggttagtcaccggagagtgattatt 240
Db 9753 ATGATGCATATTGCTAAAGATTACAGTTGGTTAGGGTTAGTCACTGGAGAGTGATTATT 9812
QY 241 ggttttttaattgggttttggcggttcccttttggcccttgatgagcggttttt 300
Db 9813 GGTTTTCAATTGGGTTTGTGGCGCGGTCCCTTTTGGGCCGTTGATATGGCGGGTTT 9872
QY 301 ctgcttgatactttacgtggcgacaaatgggttacgataattcattcacaatagaagct 360
Db 9873 CTGCTTGATACTTTACGTGGCGGCAATGGGTACGATATTCAATTTACAATAGAAGCT 9932
QY 361 gaaacctcacttttggctgttttcacccagttttgtgtgtatttttttataagc 420
Db 9933 GAAACCTCACCTTTTGGCTTGTCTTTTTCAGCCAGTTCTTGTGTATTATTCTTTATAAGC 9992
QY 421 ggcggcatggagtttatataaaacattctcgtatgagtcatacaatattaccaccaggg 480
Db 9993 GCGGCGATGGAGTTTATATTAAACATTCTGTATGAGTCATATCAATATTACCACCAGG 10052
QY 481 cgtactttattattgaccggcaatttttataaaatatatccaggcagagtggaacgctt 540
Db 10053 CGTACTTTATTATTGACCGCAATTTTAAATATATATCCAGGCAGAGTGGAGAACGCTT 10112
QY 541 tatcaattatgttcagtttctctctcctcctgccataatgatggtattagccagatctg 600
Db 10113 TATCAATTATGTATACGCTTCTCTCTCTCTGCCCAATAATATGATATGATAGCCGATCTG 10172
QY 601 gctttagggtcttttaaatcggtcggcacaaatgaatggtgttttctctcaatgccc 660
Db 10173 GCTTTAGGTCTTTTAAATCGGTGGCACAACAAATTGAATGTGTGTCTTCTCAATGGCG 10232
QY 661 ctcaaaagatatattggttctactgacgctcctgactcctcattcccttatgctcttcac 720
Db 10233 CTCAAAAGTATATTGGTTCTACTGAGC-YCCTGATCTCATTCCTTATGCTCTTCATCAC 10291
QY 721 tatttggttgaaagcagataaatttttatatttatcaaaagactgggtttccatctgtatg 779
Db 10292 TATTTCGTTGAAAGCGATAAATTTTATATTATCTCTAAAGACTGTGTTTCCATCTGTATG 10350

RESULT 10
US-08-637-759-15/c
Sequence 15, Application US/08637759
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center, 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/637,759
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: rpms 101
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Partial sequence of Salmonella typhimurium
ORGANISM: virulence gene
US-08-637-759-15

Query Match 26.6%; Score 207.4; DB 10; Length 297;
Best Local Similarity 97.2%; Pred. No. 6.8e-42;
Matches 211; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atggcacacaggtataatgagtgcttattgcatggcttgctggtgctttattcgaccatt 60
Db 224 ATGGCACACAGGTAATGAGTGGCTTATTGCAATGGCTGTGGCTTTTATTGACCATTTG 165
QY 61 agcctttcttattacttccctattataaaagtgccagtttagggcgctcttttacgt 120
Db 164 AGCCTTCTTATTACTTCCCTTATTAAAGTGGCAGTTTAGGGCGGCACACTTTTACGT 105
QY 121 aatggcggtgttatgtcacttacccttcccatattaccatatttaccagcagaagatt 180
Db 104 AATGGCGTGTATGTCACTTACCTTTCCGATATTACCAATCATTTACCAGCAGAAGATT 45
QY 181 atgatgcataattggttaaaagattcacagttcgttaggtt 217
Db 44 ATGATGCATATTGGTAAAGATTACAGTTGGGTAGGTT 8

RESULT 11
US-08-637-759A-15/c


```
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: Novel Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 206:
SEQUENCE CHARACTERISTICS:
LENGTH: 43360
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 206:
US-09-453-702B-206

Query Match 15.1%; Score 118; DB 18; Length 43360;
Best Local Similarity 52.0%; Pred. No. 1.2e-18;
Matches 265; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

Qy 208 tggtagggtagtcaccgagaggtgattatgggttttttaattgggttttgctgcgcgcg 267
Db 35984 TTGACAGGAATAGCGCTAAAGAAATAAGTATGGTGTTCATTGGGTATCATTTACT 35925

Qy 268 gtcccttttgggcccgttgatagcgggggtttctgctgatactttacgtgcccgcgaca 327
Db 35924 ATTCTTTTGGGCAATAGATCGCGCTGCACAGATTATTGATACTCTAAGAGGCTCAACA 35865

Qy 328 atggglacgatattcaattctacaatagaagctgaacctcacttttgggttgccttttc 387
Db 35864 ATATCTCAATTTTAACCGCTCCATAGTATCATCTTCTATCATCTGGCGTTATCTTG 35805

Qy 388 agccagtttttgggtgtattttttttaaagcggcgccgaggttttataataaacatt 447
Db 35804 TACCAATTTATCTCTGTGATCTTTGTTTATTCATGGTGGGATACAAAGCATCTAGATAAG 35745

Qy 448 ctgtatgagtcataatatttaccaccaggcggtactttattattttgacccggcaattt 507
Db 35744 CTATATTTATCTACGAGATATTACCATTTACAAGCCGATATTGCATTTCAATCGTGCTTTA 35685

Qy 508 ttaaaatataatccaggcagagtgagagaacgctttatcaattatggtcagttctctctt 567
Db 35684 ATAGATTTTTTGTCTCTATGGGATTCATTATTAAACTGATGTTATCATTTTCAGTT 35625

Qy 568 cctggcataatagtattgattagccgatctggccttttggttcttttaaaatgggtcgcca 627
Db 35624 CCCATGATTATCGGTATATCTTATGTGATATGGGTTTGGGTTTCTTTTAAACAAACAGCA 35565

Qy 628 caacaattgaatgtgtttttctctcaatgcctcctcaaaagtatatgttctactgcgcg 687
Db 35564 CCTCAGCTAAAGTATTCACATTATCCTACCTGCCAGTAAAGTTTGTATAGCAATCTTTATA 35505

Qy 688 ctccgtatctcattcccttattgctcttcat 717
Db 35504 TTGCTATTAGTATTATTCATGTTTTCCTGAT 35475

RESULT 14
US-09-453-702B-261/C
; Sequence 261, Application US/09453702B
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: Novel Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 261:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 261:
US-09-453-702B-261

Query Match 15.1%; Score 118; DB 18; Length 45325;
Best Local Similarity 52.0%; Pred. No. 1.2e-18;
Matches 265; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

Qy 208 tggtaggggttagtcaccgagagaggtgattatgggttttttaattgggttttgctgcgcgcg 267
Db 36886 TTGACAGGAATAGCGCTAAAGAAATAAGTATGGTGTTCATTGGGTATCATTTACT 36827

Qy 268 gtcccttttgggcccgttgatagcgggggtttctgctgatactttacgtgcccgcgaca 327
Db 36826 ATTCCTTTTGGGCAATAGATCGCGCTGCACAGATTATTGATACTCTAAGAGGCTCAACA 36767

Qy 328 atggggtagcatatttcaattctacaatagaagctgaacctcacttttgggttgccttttc 387
Db 36766 ATATCTCAATTTTAAACCGCTCCATAGTATTCATCTTCTATCATCTGGCGTTATCTTG 36707
```

```
QY 388 agcaggttttgggtggttatttcttataagcggcgagtgagtttataataacatt 447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36706 TACCAATTTATCTCTGTGATCTTTGTTATTCATGTTGGGGATACAAAGCATCCTAGATAAG 36647

QY 448 ctgtatgagtcataatattaccaccaggcgctactttattatttgaccggcaattt 507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36646 CTATATTTATCTACGAGATATACCAATTACAAGCCGATATTCGATTCATTCGTCCTTA 36587

QY 508 ttaaaatataatccaggcagagtgagagacgctttatcaattatggtcagtttctctt 567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36586 ATAGATTTTGTGTTTCTCTATGGGATTCATTTATTAACCTGATGTTATCATTTTCAGTT 36527

QY 568 cctgccataatattggttattagcagatctggttttaggtcttttaaatcggtcgga 627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36526 CCCATGATATTCGGTATATCTTATGATGATGTTGGGTTTGGGTTTCTTAACAAACAGCA 36467

QY 628 caacaattgaatggttttttcttcaatgcgcgtcaaaagtattggttctactgacg 687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36466 CCTCAGCTAAACGTATTCACATTTACACTGCCAGTAAAGTTTGTATAGCAATCTTTATA 36407

QY 688 ctcctgatctcattcccttatgctcttcat 717
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36406 TTGCTATTAGTTATTCATGTTTTCCTGAT 36377
```

RESULT 15

```
US-09-409-800B-1/c
; Sequence 1, Application US/09409800B
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Burland, Valerie
; APPLICANT: Rose, Debra J.
; APPLICANT: Mayhew, George F.
; APPLICANT: Perna, Nicole
; APPLICANT: Perry, Robert D.
; APPLICANT: Straley, Susan C.
; APPLICANT: Fetherston, Jacqueline D.
; APPLICANT: Lindler, Luther E.
; APPLICANT: Plano, Gregory V.
; TITLE OF INVENTION: Plasmid DNA From Yersinia Pestis
; FILE REFERENCE: 960296.95939
; CURRENT APPLICATION NUMBER: US/09/409,800B
; CURRENT FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 70559
; TYPE: DNA
; ORGANISM: Yersinia pestis
US-09-409-800B-1
```

```
Query Match          9.8%; Score 76; DB 18; Length 70559;
Best Local Similarity 48.2%; Pred. No. 5.6e-08;
Matches 245; Conservative 0; Mismatches 260; Indels 3; Gaps 1;

QY 193 ggtaaagattacagttggttaggttagtcacccggagaggtgatttggtttttaatt 252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27477 GTTGATGCGTTTACGTTGATGCTGCTTATCGGAAAGAGATCATACTGGGGTTATTGATT 27418

QY 253 ggggttttggcggttcccttttggcggttgatagtcgggggtttctgctgatact 312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27417 GGGTTTGGCCACCATTCTCTTCTGGGCCCTTAGAATCCGAGGATTATTGTAGATAAC 27358

QY 313 ttacgtggcgagacaatgggtacgatattcaattctacataagactgaacctcactt 372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27357 CAAGAGGTGCGCGATGGCATCTCTACTCAATCTGGACTTGATAGTCAAACTAGTCCG 27298

QY 373 ttggcttgccttttcagccagtttttgggtgtgttatttttttaagcgcgcatggag 432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27297 ACTGGTTTACTTTTGGACTCAACCGTTAATAACAAATTTTCTTACGGGGGGGCTTCTC 27238

QY 433 ttatatataacattctgtatgagtcataatattaccaccaggcgctactttatta 492
```

```
Db 27237 TCCTTGTCTTCAGCCCTCTTTTCACAGCTATGTAAATTTGGCCGTTGGCCAGCTTTTTCCTT 27178

QY 493 ttgaccgggcaatttttaaaataataatccaggcagagtgagagacgcttttcaattatgt 552
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27177 GCAGTTAGTGAACAGTGGGTTTCATTTCTTTTATAACCAATTTCAGCCAGATACATTAATC 27118

QY 553 gtcagtttctcttctcctcctccataataatgtatggtattagcagatctggccttaggt 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27117 GCGGCTGTATTGGCTGCTCTTTACTCATCGCTATGTTTTCAGCTGAATTTGGACTTGCA 27058

QY 610 cttttaaatcggtcggcgacacaaattgaattggtgttttctctcctcaatgccgctcaaaagt 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27057 CTCATCAGTCGTTTTCGGCCCTCTTTTAAACGCTCTTTGTGCTGGCTATGCCGATAAAAAGC 26998

QY 670 atattggttctactgacgctcctgact 697
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 26997 GCGATACCAAGCCTGTTGTTGGTTATCT 26970
```

Search completed: August 2, 2002, 03:50:14
Job time: 7579 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 02:16:20 ; Search time 4280.15 Seconds
(without alignments)
395.143 Million cell updates/sec

Title: US-09-809-524B-1
Perfect score: 779
Sequence: 1 atggcacaacagtgtaataa.....gactgggttcacatctgtatg 779

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1359166 seqs, 1085540162 residues

Total number of hits satisfying chosen parameters: 2718332

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New.*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	118	15.1	43360	6 US-10-114-170-206	Sequence 206, App
C 2	118	15.1	45325	6 US-10-114-170-261	Sequence 261, App
C 3	53.6	6.9	16950	6 US-10-114-170-166	Sequence 166, App
C 4	37.2	4.8	14147	7 US-10-172-096-51	Sequence 51, Appl
C 5	36.6	4.7	283	5 US-09-539-800C-12433	Sequence 12433, A
C 6	36.6	4.7	301	5 US-09-539-800C-14817	Sequence 14817, A
C 7	36.6	4.7	22713	7 US-10-105-239-12732	Sequence 12732, A
C 8	36	4.6	1145	8 US-60-391-781-423	Sequence 423, App
9	35.2	4.5	621	7 US-10-027-632-313321	Sequence 313321, A
10	35.2	4.5	621	7 US-10-027-632-313322	Sequence 313322, A
11	35.2	4.5	2460	7 US-10-104-047-770	Sequence 770, App
12	35	4.5	536	5 US-09-721-544-1708	Sequence 1708, App
C 13	34.8	4.5	168575	7 US-10-178-194-1	Sequence 1, Appl
14	34.4	4.4	509	7 US-10-027-632-87445	Sequence 87445, A
15	34.2	4.4	611	7 US-10-027-632-179972	Sequence 179972, A
C 16	34.2	4.4	647	7 US-10-027-632-187203	Sequence 187203, A
17	34.2	4.4	2484	5 US-09-629-469A-17966	Sequence 17966, A
18	34.2	4.4	2713	5 US-09-629-469A-16837	Sequence 16837, A
C 19	34	4.4	520	5 US-09-918-995-19426	Sequence 19426, A
20	34	4.4	98666	6 US-10-189-267-20	Sequence 20, Appl
C 21	33.8	4.3	465	5 US-09-918-995-10260	Sequence 10260, A
C 22	33.8	4.3	1295	7 US-10-144-132-33	Sequence 33, Appl
C 23	33.8	4.3	113585	6 US-10-188-470-32	Sequence 12, Appl
C 24	33.6	4.3	311	5 US-09-991-936-1266	Sequence 1266, App
C 25	33.6	4.3	585	7 US-10-027-632-193565	Sequence 193565, A

26	33.6	4.3	585	7 US-10-027-632-235108	Sequence 235108, A
27	33.6	4.3	585	7 US-10-027-632-235109	Sequence 235109, A
28	33.6	4.3	585	7 US-10-027-632-235110	Sequence 235110, A
29	33.6	4.3	585	7 US-10-027-632-235111	Sequence 235111, A
30	33.6	4.3	585	7 US-10-027-632-235112	Sequence 235112, A
C 31	33.6	4.3	660	7 US-10-027-632-248133	Sequence 248133, A
C 32	33.6	4.3	676	7 US-10-027-632-224256	Sequence 224256, A
33	33.4	4.3	767	7 US-10-027-632-18755	Sequence 18755, A
34	33.4	4.3	767	7 US-10-027-632-18756	Sequence 18756, A
C 35	33.4	4.3	1936	7 US-10-105-239-28556	Sequence 28556, App
36	33.2	4.3	357	5 US-09-721-544-12126	Sequence 12126, A
C 37	33.2	4.3	507	7 US-10-027-632-223574	Sequence 223574, A
38	33.2	4.3	512	8 US-60-377-240-7220	Sequence 7220, App
39	33.2	4.3	520	7 US-10-027-632-34696	Sequence 34696, A
40	33.2	4.3	520	7 US-10-027-632-34697	Sequence 34697, A
C 41	33.2	4.3	728	7 US-10-027-632-145891	Sequence 145891, A
42	33.2	4.3	735	7 US-10-027-632-27598	Sequence 27598, A
C 43	33.2	4.3	1462	1 PCT-US02-19457-163	Sequence 163, App
C 44	33.2	4.3	1462	7 US-10-175-523-163	Sequence 163, App
45	33.2	4.3	1473	5 US-09-540-209B-129	Sequence 129, App

ALIGNMENTS

RESULT 1
US-10-114-170-206/c
; Sequence 206, Application US/10114170
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: Novel Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 206:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43360
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLFCULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 206:
US-10-114-170-206

Query Match 15.1%; Score 118; DB 6; Length 43360;
Best Local Similarity 52.0%; Pred. No. 1.8e-19;
Matches 265; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

QY 208 tgggttaggttagtcacccgagaggtgattattggtttttaaattggttttgcgagcg 267
Db 35984 TTGACAGGAATAGCGCTAAAGAAATAGATTGGTTTTCATTTGGGTATCATTTACT 35925

QY 268 gtcccttttgggcccgttgatgagcggggttctgctgatacttaccgtgagcgagaca 327
Db 35924 ATTCTTTTGGGCAATAGATGCGGTGGACAGATTATGATCTCTAGAGGCTCAACA 35865

QY 328 atgggttagcgaattcaattctacaatagagctgaagcctcaacttttggcgttcttcc 387
Db 35864 ATATCTTCAATTTTAAACCGCTCCATAAGTATTCATCTTATCACCTGGCGTTATCTTG 35805

QY 388 agccagttttgtgtgtatttcttcttataagcgcgagcgtgaggtttatataaacatt 447
Db 35804 TACCAATTTATCTCTGTGATCTTTGTTTATTCATGTTGGGATACAAAGCATCTAGATAAG 35745

QY 448 ctgtatgagtcataatcaatttaccacagggcggtactttatttatttggacggcaattt 507
Db 35744 CTATATTTATCTACGAGATATTACCATTTACAGCGGATATTGCATTCATTCGCTTTA 35685

QY 508 ttaaaatataatccagggagagtgagagacgctttatcaattatgtcagtttctctctt 567
Db 35684 ATAGATTTTGTGTTCTCTATGGGATTCATTTATTAACCTGATGTTTATCATTTTCAGTT 35625

QY 568 cctgcaataatgattgattagcgcgctgcttcttcttcttcttcttcttcttcttctt 627
Db 35624 CCCATGATTATCGGTATATCTTATGTGATATGGGTTGGGTTTCTTAAACAAACAGCA 35565

QY 628 caacaattgaatggttttcttcttcttcttcttcttcttcttcttcttcttcttctt 687
Db 35564 CCTCAGCTAAACGATTTCATATTCATCTCCAGTAAAGTTTGTAGAGCAATCTTTATA 35505

QY 688 ctctgatctcatctcccttcttcttcttcttcttcttcttcttcttcttcttctt 717
Db 35504 TTGCTATTAGTTATTCATGTTTTTCCGTAT 35475

RESULT 2

US-10-114-170-261/c

; Sequence 261, Application US/10114170

; GENERAL INFORMATION:

; APPLICANT: Blattner, Frederick R.

; Burland, Valerie

; Plunkett, Guy

; Welch, Rod

; TITLE OF INVENTION: Novel Sequences of E. coli O157

; NUMBER OF SEQUENCES: 265

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Quarles & Brady

; STREET: 1 South Pinckney Street

; CITY: Madison

; STATE: WI

; COUNTRY: US

; ZIP: 53701-2113

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 8.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/114,170

; FILING DATE: 01-Apr-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/453,702

; FILING DATE: 03-DEC-1999

APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 261:
SEQUENCE CHARACTERISTICS:
LENGTH: 45325 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 261:
US-10-114-170-261

Query Match 15.1%; Score 118; DB 6; Length 45325;
Best Local Similarity 52.0%; Pred. No. 1.9e-19;
Matches 265; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

QY 208 tgggttaggttagtcacccgagaggtgattattggtttttaaattggttttgcgagcg 267
Db 36886 TTGACAGGAATAGCGCTAAAGAAATAGATTGGTTTTCATTTGGGTATCATTTACT 36827

QY 268 gtcccttttgggcccgttgatgagcggggttctgctgatacttaccgtgagcgagaca 327
Db 36826 ATTCTTTTGGGCAATAGATGCGGTGGACAGATTATGATCTCTATCACCTGGCGTTATCTTG 36707

QY 328 atgggttagcgaattcaattctacaatagagcgtgaggttttggcgttcttcttcttctt 387
Db 36766 ATATCTTCAATTTTAAACCGCTCCATAAGTATTCATCTTCTATCACCTGGCGTTATCTTG 36707

QY 388 agccagttttgtgtgtatttcttcttataagcgcgagcgtgaggtttatataaacatt 447
Db 36706 TACCAATTTATCTCTGTGATCTTTGTTTATTCATGTTGGGATACAAAGCATCTAGATAAG 36647

QY 448 ctgtatgagtcataatcaatttaccacagggcggtactttatttatttggacggcaattt 507
Db 36646 CTATATTTATCTACGAGATATTACCATTTACAGCGGATATTGCATTCATTCGCTTTA 36587

QY 508 ttaaaatataatccagggagagtgagagacgctttatcaattatgtcagtttctctctt 567
Db 36586 ATAGATTTTGTGTTCTCTATGGGATTCATTTATTAACCTGATGTTTATCATTTTCAGTT 36527

QY 568 cctgcaataatgattgattagcgcgctgcttcttcttcttcttcttcttcttcttctt 627
Db 36526 CCCATGATTATCGGTATATCTTATGTGATATGGGTTTGGGTTTCTTAAACAAACAGCA 36467

QY 628 caacaattgaatggttttcttcttcttcttcttcttcttcttcttcttcttcttctt 687
Db 36466 CCTCAGCTAAACGATTATTCATATTCATCTCCAGTAAAGTTTGTAGATCAACAACTTTATA 36407

QY 688 ctctgatctcatctcccttcttcttcttcttcttcttcttcttcttcttcttctt 717
Db 36406 TTGCTATTAGTTATTCATGTTTTTCCGTAT 36377

RESULT 3

US-10-114-170-166/c

; Sequence 166, Application US/10114170

; GENERAL INFORMATION:

; APPLICANT: Blattner, Frederick R.

; Burland, Valerie

; Perna, Nicole T.

; Plunkett, Guy

; Welch, Rod

; TITLE OF INVENTION: Novel Sequences of E. coli O157

; NUMBER OF SEQUENCES: 265

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/114,170
FILING DATE: 01-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/453,702
FILING DATE: 03-DEC-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 166:
SEQUENCE CHARACTERISTICS:
LENGTH: 16950
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 166:
US-10-114-170-166

	Query Match	6.9%	Score 53.6;	DB 6;	Length 16950;
	Best Local Similarity	49.6%;	Pred. No. 0.0015;		
	Matches 137;	Conservative	0;	Mismatches 139;	Indels 0; Gaps 0;
QY	216	gttagtcacgcggagagggtgattatttggtttttaaattgggttttgtcgccgggttcacct	275		
Db	5518	GTTAGCTGCAGAGAATATTATCATAGGGCGTTCATTGCTTGCTTTATTAGCATCTCCATT	5459		
QY	276	ttaggcgcttgatatggcgggggtttctgctgtgatactttacgtggcgcgcaaatgggtac	335		
Db	5458	TTCGATTTTTCTTGCTATAGTAGTACGTATTATCGATAACCACCGGTGCAACGTTAAGTAG	5399		
QY	336	gatattccaattctacaatagaagctgaacacctcacctttttggcttgcttttcacgccagtt	395		
Db	5398	TACTTTGGATCCGGCTACAGGTGTCGATACCTCGGAATTAGCAAGATTATTAAACCTTTT	5339		
QY	396	tttgtgtgtattttctttataagcggcggcatggagtttatattaaacattctgtatga	455		
Db	5338	CTCTGCAGCGGTTTATCTTTACTAACGGTGGATTGAACCTTTATTCTGGAAACTCTTTAGCA	5279		
QY	456	gtcataatcaattattaccaccaggcggctactttatt	491		
Db	5278	AAGTTTAATTTTATGGCCTTCAGGAAGCTTTTAATTT	5243		

```

RESULT      4
US-10-172-086-51
; Sequence 51, Application US/10172086
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Method and nucleic acids for the differentiation
; TITLE OF INVENTION: of prostate tumors
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/172,086
; CURRENT FILING DATE: 2002-06-13

```

[illegible]

```

RESULT      5
US-09-539-800C-12433/c
; Sequence 12433, Application US/09539800C
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF CONNECTIVE TISSUE
; FILE REFERENCE: PD-1023 CIP
; CURRENT APPLICATION NUMBER: US/09/539,800C
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 08/521,383
; PRIOR FILING DATE: August 16, 1995
; PRIOR APPLICATION NUMBER: 08/271,217
; PRIOR FILING DATE: June 27, 1994
; PRIOR APPLICATION NUMBER: 08/334,881
; PRIOR FILING DATE: November 4, 1994
; PRIOR APPLICATION NUMBER: 08/943,978
; PRIOR FILING DATE: October 3, 1997
; PRIOR APPLICATION NUMBER: 60/028,732
; PRIOR FILING DATE: October 4, 1996
; PRIOR APPLICATION NUMBER: 08/943,979
; PRIOR FILING DATE: October 4, 1997
; PRIOR APPLICATION NUMBER: 60/027,782
; PRIOR FILING DATE: October 4, 1996
; PRIOR APPLICATION NUMBER: 08/993,774
; PRIOR FILING DATE: December 18, 1997
; PRIOR APPLICATION NUMBER: 60/034,975
; PRIOR FILING DATE: December 20, 1996
; PRIOR APPLICATION NUMBER: 09/250,003
; PRIOR FILING DATE: February 10, 1999

```


RESULT 11

```
US-10-104-047-770
; Sequence 770, Application US/10104047
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 770
; LENGTH: 2460
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-770

Query Match          4.5%; Score 35.2; DB 7; Length 2460;
Best Local Similarity 46.4%; Pred. No. 40;
Matches 115; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
```

```
QY 233 tgattattggttttttaattgggttttgcggcggttccttttggcggttgatgg 292
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2092 tgatgcdaattcttctaatttttttgatgttggtgtcttttccaggagtgagaatct 2151
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 293 cgggggtttctgtgatactttacgtggcgcgacaatgggtacgatatccaattctacaa 352
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2152 atgtattggtggtgtcattttccagctctctcgtgcattcttggtagaattctaac 2211
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 353 tagaagctgaacctcactttttggttgccttttcagccagtttttgggttattttct 412
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2212 tagagatgatacattctttaaactgttggtatccatgcttctgtctgtacatga 2271
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 413 ttataagcggcgcatgagtttatattaaacattctgtatgagtcataatctttac 472
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2272 ccattgttcttgacatggaataaccagtcagtggtctctatgtattgtctatatgtaa 2331
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 473 caccagg 480
||| ||| |||
Db 2332 accaagag 2339
||| ||| |||
```

RESULT 12

```
US-09-721-544-1708
; Sequence 1708, Application US/09721544
; GENERAL INFORMATION:
; APPLICANT: Arterburn, Matthew
; APPLICANT: Asghari, Vida
; APPLICANT: Damavandi, Simin
; APPLICANT: Dickson, Mark
; APPLICANT: Drake, Jim
; APPLICANT: Drmanac, Radoje
; APPLICANT: Engleman, Carrie
; APPLICANT: Faulkner, Brandy
; APPLICANT: Garcia, Veronica
; APPLICANT: Gledt, Gretchen
; APPLICANT: Hunter, Kelly
; APPLICANT: Jessen, Aaron
; APPLICANT: Jones, Lee
; APPLICANT: Kita, David
; APPLICANT: Labat, Ivan
; APPLICANT: Laroya, Mimi
; APPLICANT: Lomelli, Michelle
; APPLICANT: Nguyen, Phuong
; APPLICANT: Nogra, Margie
; APPLICANT: Palencia, Servando
; APPLICANT: Ralsi, Fariba
; APPLICANT: Smith, Benjamin
; APPLICANT: Tkach, Joe
; APPLICANT: Tran, Lien
```

```
; APPLICANT: Verna, Ron
; APPLICANT: Yang, Fei
; APPLICANT: Yim, Kenneth
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
; FROM A cDNA Library of Fetal Liver-Spleen
; FILE REFERENCE: 728CIP
; CURRENT APPLICATION NUMBER: US/09/721,544
; CURRENT FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 09/515,128
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: 09/034,341
; PRIOR FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 24489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1708
; LENGTH: 536
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(536)
; OTHER INFORMATION: n = A,T,C or G
US-09-721-544-1708
```

```
Query Match          4.5%; Score 35; DB 5; Length 536;
Best Local Similarity 54.2%; Pred. No. 33;
Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
```

```
QY 371 ttttgggttgccttttcagccagtttttgcgttattttttataagcggcgcgcatgg 430
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 387 ttctgggtgagcttttaagtcagtttttgcgttatttttgcgttatttttaagtttgaaataa 446
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 431 agttattataaacattctgtatgagtcataatattaccaccagggcggtactttat 490
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 447 actcattagcaaacacttagaataatccaaagtaataagatagatatttttggcaccagtttt 506
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 491 tatttgaccgg 501
||| ||| |||
Db 507 gttttgtatgg 517
||| ||| |||
```

RESULT 13

```
US-10-178-194-1/c
; Sequence 1, Application US/10178194
; GENERAL INFORMATION:
; APPLICANT: Berglund Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/10/178,194
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US/09/426,290
; PRIOR FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
US-10-178-194-1

Query Match      4.4%; Score 34.8; DB 7; Length 168575;
Best Local Similarity 52.8%; Pred. No. 1.1e+02;
Matches 75; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 427 atggaggttattataacattctgtatgagtcataatcaattttaccaccaggcggtact 486
Db 18358 ATTGAAGTTATAAATAACGAGAAATGAAATATTTTCAAGATTTTATCCTGATACCTGCT 18299

QY 487 ttattattgacggcgaatttttaaaatatatccaggcagagtgagagaacgctttatcaa 546
Db 18298 TTATTTTATGACAAAGCGTATTATCTTGATCCTACAAATATTCATTTCTCTATCTA 18239

QY 547 ttatgtgtcagttctctcttc 568
Db 18238 TGATTTCTCTGGTCTTCTTC 18217

RESULT 14
US-10-027-632-87445
; Sequence 87445, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 87445
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-87445
```

```
Query Match      4.4%; Score 34.4; DB 7; Length 509;
Best Local Similarity 46.2%; Pred. No. 46;
Matches 110; Conservative 1; Mismatches 127; Indels 0; Gaps 0;

QY 336 gataattcaattctacaagaagctgaacacctcaatttttggctgtcttttcagccagtt 395
Db 51 gatcttaagtgaatgatgtatacaaaaactacatttttccacagcattataagaat 110

QY 396 ttatgtgttattttctataagcggcgagtgagtttatataaacattctgtatga 455
Db 111 gttgaaygaacaatttttttgaggagacttggtatatatagtttttgcttaaaagctgtttcc 170
```

```
QY 456 gtcatatcaaatattaccaccaggggcgctactttattattttaccggcgaatttttaaaaa 515
Db 171 aagaacttatgatgacatttaaggactgtctatatctgtcccttacttccatcttc 230

QY 516 tatccaggcagagtgagacgctttatcaattatgtcagtttctctctctccgcc 573
Db 231 tctgtaagtgtactctcacaaactatttctccagtcctcaaatatctgtgaaatatctgcc 288

RESULT 15
US-10-027-632-179972
; Sequence 179972, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 179972
; LENGTH: 611
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-179972
```

```
Query Match      4.4%; Score 34.2; DB 7; Length 611;
Best Local Similarity 60.0%; Pred. No. 54;
Matches 57; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 111 tctttacgtaatgctgtcttatgtcacttaccctttcccatattaccatattacca 170
Db 205 tatatcccatatttagcctcaatttttcaattttcccttattatgtttccattttatca 264

QY 171 gcagaagattatgatgcataattggtaaagattaca 205
Db 265 gaataagattattttttatgtaggcaaacatttcca 299
```

Search completed: August 2, 2002, 05:02:21
Job time: 9961 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 00:37:05 : Search time 1811.32 Seconds
(without alignments)
8999.940 Million cell updates/sec

Title: US-09-809-524b-1
Perfect score: 779
Sequence: 1 atggcacaacaggtaaatga.....gactgggtttccatctgtatg 779

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_be:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	779	100.0	779	6	AX254677
2	763	97.9	779	6	AX254678
3	763	97.9	5484	1	STYSCQRS
4	763	97.9	27863	1	AE008761
5	750.6	96.4	13417	6	AS1693
6	750.6	96.4	13417	6	AX001627
7	207.4	26.6	297	6	AS1671
8	207.4	26.6	297	6	AX001605
9	180.8	23.2	288050	1	AJ414141
10	119.6	15.4	35624	1	AF022236
11	118	15.1	11760	1	AE005597
12	118	15.1	45325	1	AF071034
13	118	15.1	280900	1	AF002566
14	113.8	14.6	37889	1	AF200363
15	113.8	14.6	60073	1	ECO27443
16	113.8	14.6	60354	1	AF453441
17	99.2	12.7	42001	1	AF311901
18	76	9.8	5752	1	YEPYSC
19	76	9.8	70305	1	YPCD1
20	76	9.8	70504	1	AF053946
21	76	9.8	70559	1	AF074612
22	73.8	9.5	67720	1	AF336309
23	70.6	9.1	110000	2	LMFLCHR32_02
24	67.8	8.7	7874	1	SHFPMYSH
25	67.8	8.7	20323	1	SHFORF
26	67.8	8.7	213494	1	SPWR100
27	67.8	8.7	221618	1	AF386526
28	67.8	8.7	221851	1	AF348706
29	66.6	8.5	303249	1	AP001515
30	66.4	8.5	69673	1	AF102990
31	60.2	7.7	801	6	AX028438
32	59.2	7.6	11958	1	AE004596
33	53.8	6.9	13008	1	AE007715
34	53.6	6.9	10875	1	AE005515
35	52.4	6.7	3350	1	AF172245
36	52.2	6.7	184	6	AS1672
37	52.2	6.7	184	6	AX001606
38	50	6.4	1641	1	AF306650
39	49.6	6.4	9613	1	AB052736
40	47.2	6.1	7218	6	I66494
41	44.6	5.7	1673	8	AF284928
42	44	5.6	124508	30	AL139017
43	44	5.6	124975	2	AC104455
44	44	5.6	197605	2	AC026905
45	43.8	5.6	117847	2	AC095778

ALIGNMENTS

RESULT 1

AX254677

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

AX254677 Sequence 1 from Patent WO0170247.
AX254677
AX254677.1 GI:16074344

Salmonella enterica subsp. enterica serovar Dublin.
Salmonella enterica subsp. enterica serovar Dublin.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.

1 (bases 1 to 779)
Lowery, D.E. and Kennedy, M.J.
Salmonella vaccine materials and methods
Patent: WO 0170247-A 1 27-SEP-2001;
Pharmacia & Upjohn (US)

Location/Qualifiers

1..779

/organism="Salmonella enterica subsp. enterica serovar

Dublin"

/db_xref="taxon:98360"

185 a 143 c 161 g 290 t

BASE COUNT

ORIGIN

JOURNAL Patent: WO 0170247-A 2 27-SEP-2001;
Pharmacia & Upjohn (US)

FEATURES Location/Qualifiers
Source 1..779
/organism="Salmonella typhimurium"
/db_xref="taxon:602"

BASE COUNT 187 a 143 c 161 g 288 t

ORIGIN

Query Match 97.9%; Score 763; DB 6; Length 779;
Best Local Similarity 98.7%; Pred. No. 8.3e-174;
Matches 769; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

QY 1 atggcacacaggtaaatgagtggtctattgcattggctgtggctttattcgaccatta 60
   |||||
Db 1 ATGGCACACAGGTAATGATGGCTTATTCGATTGGCTGTGGCTTTATTTCGACCATG 60

QY 61 agcctttctttattacttccccattaaaaagtgccagtttagggccgcctctttta 120
   |||||
Db 61 AGCCTTCTTTTATTACTTCCCTATTAAAAAGTGGCAGTTTAGGGCCGCACTTTTACGT 120

QY 121 aatggcgtgcttatgtcacttacccttccccattaccattaccattaccagcagaagt 180
   |||||
Db 121 AATGGCGTGCTTATGTACCTTACCTTCCGATATTACCAATCATTTTACCAGCAGAAGATT 180

QY 181 atgatgcataattggtaaaagattacagtttgtaggttagtcaccggagaggtgattatt 240
   |||||
Db 181 ATGATGCATATTGGTAAAGATTACAGTTGGTTAGGTTAGTCACTGGAGAGGTGATTATT 240

QY 241 gctttttaattgggttttggcggggttcccttttggccgttgatggcggggtttt 300
   |||||
Db 241 GGTTTTCAATTGGGTTTGTGCGCGGTTCCCTTTTGGCGCGTTGATATGGCGGGGTTT 300

QY 301 ctgcttgatactttaogtggcgcgacaaatgggtacgatatccaattctacaataagaagct 360
   |||||
Db 301 CTGCTTGATACCTTACGTGGCGCGCAATGGGTAGCGATATTCAATTCTACAATAGAAGCT 360

QY 361 gaaacctcaacttttggcttcttttccagccagtttttgctgtgtatttcttttataagc 420
   |||||
Db 361 GAAACCTCACTTTGTGGCTTCTTTCAGCCAGTCTCTGTGTGTATTTCCTTTTATAAGC 420

QY 421 ggcggcaatgggtttattataaacattctgtatgagtcataatattaccaccagg 480
   |||||
Db 421 GCGGCATGGAGTTTATATTAACATCTCTGTATGAGTCATATCAATATTACCACAGGG 480

QY 481 cgtacttattattgaccggcaatttttaaaatataatccaggcagagtggagaaacgctt 540
   |||||
Db 481 CGTACTTTATTATTGACCACGAATTTTAAAAATATATCCAGGCACAGTGGAGAACGCTT 540

QY 541 tatcaattatgtcagtttctctcttccgtgccataaatgtatggtattagcogattc 600
   |||||
Db 541 TATCAATTATGATACAGTCTCTCTCTCCGCAATAATATGATATGGTATTAGCGGATCTG 600

QY 601 gctttaggtcttttaaatcggtcgggcacacaattgaatgtgttttcttctcaatgcg 660
   |||||
Db 601 GCTTTAGGTCCTTTAAATCGGTCGGCACACAAATTGAATGTGTTTTCTTCTCAATGCCG 660

QY 661 ctcaaaagtatatgtttctactgaagcctctgatctcaattcccttatgctcttctatcac 720
   |||||
Db 661 CTCAAAGTATATTGGTTCTACTGACGCTCTGATCTCATTCCTTATGCTCTTCATCAC 720

QY 721 tatttgggtgaagcgataaattttattattatactaaaagactgggttccatctgtatg 779
   |||||
Db 721 TATTTGGTTGAAGCGGATAAATTTTATATTTATCTAAAGACTGGTTTCCATCTGTATG 779

```

RESULT 3
STYSCQS
LOCUS STYSCQS 5484 bp DNA linear BCT 28-FEB-1997
DEFINITION S.typhimurium ssaQ, ssaR, ssaS, ssaT and ssaU genes.
ACCESSION X99944
VERSION X99944.1 GI:1526983

Query Match 100.0%; Score 779; DB 6; Length 779;
Best Local Similarity 100.0%; Pred. No. 1.2e-177;
Matches 779; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcacacaggtataatgagtggcttattgcattgctgtgctgttttatttcaccaccatta 60
|||||
DB 1 ATGGCACACAGGTAAATGAGTGGCTTATTGCATTGGCTGTGGCTTTTATTTCACCATTTA 60
|||||

QY 61 agcctttctttattacttccccctattaaaaaggcgagtttaggggcccgcctcttttacgt 120
|||||
DB 61 AGCCTTTCTTTATTACTTCCCTTATTAAAAAGTGCGACTTAGGGGCCGCTCTTTTACGT 120
|||||

QY 121 aatggcgctgttatgtcaacttaacctttcccatattaccaaatctttaccagcagaagatt 180
|||||
DB 121 AATGGCGTGTCTTATGTCACTTACCTTTCCCATATTACCAATCATTTACCAGCAGAAGATT 180
|||||

QY 181 atgatgcattatgttgtaaagattacagtttggttagtgaccgcgagaggtgtattatt 240
|||||
DB 181 ATGATGCATATTGGTAAAGATTACAGTTGGTTAGGTTAGTGTACCGGAGAGGTGATTATT 240
|||||

QY 241 ggttttttaattgggttttgtgcggcggttcccttttggggccgctgtatatggcggggttt 300
|||||
DB 241 GGTTTTTTAATTGGGTTTGTGCGCGGTTCCCTTTTGGCGCTTGATATGGCGGGGT 300
|||||

QY 301 ctgcttgatactttacgtggcgcgacaatgggtacgatatccaattctacaaatagaaqct 360
|||||
DB 301 CTGCTTGATACTTTACGTGGCGGACAATGGGTACGATATCCAATCTCATAATAGAACT 360
|||||

QY 361 gaacacctacttttggctgttttccagccagtttttgggtgttttctttttaaacg 420
|||||
DB 361 GAAACCTCACTTTTGGCTTGTCTTTCAGCCAGTTTGTGTGTATTTCCTTTATAAGC 420
|||||

QY 421 ggcgcatggagtttatataaacattctgtatgagtcataatcaaatattaccaccagg 480
|||||
DB 421 GCGGGCATGGAGTTTATATAAACATCTGTATGAGTCATATCAATATTACCACCAGG 480
|||||

QY 481 cgtactttattattggaccggcaatttttaaaaatatccaggcagagtggaagacgctt 540
|||||
DB 481 CGTACTTTATTATTGACCGGCAATTTTAAAATATATCCAGGCAGAGTGGAGAACGCTT 540
|||||

QY 541 tatcaattatgtctcaattctctctcttcccgcataatatgtatggtattagccgactcg 600
|||||
DB 541 TATCAATTATGTCTAGTTCTCTCTCTCTCTGTCATATATGTATGTATAGCCGATCTG 600
|||||

QY 601 gctttagggtcttttaaatcgctcggcacaaattgaatglttttcttctcaatgcg 660
|||||
DB 601 GCTTTAGTCTTTTAAATCGGTGGCACAACAATGAATGTGTCTTTCTCAATGCCG 660
|||||

QY 661 ctcaaagatattggttctactgacgctcctgatctcattcccttatgtcttctcac 720
|||||
DB 661 CTCAAAAGTATATTGGTTCTACTGACGCTCCTGATCTCATTTCCCCTTATGCTCTTCATC 720
|||||

QY 721 tatttggttgaaagcagataaatttttatatttctaaaaagactggtttccatctgtatg 779
|||||
DB 721 TATTGGTTGAAAGCGATAAATTTTATATTATCTAAAAGACTGGTTTCCATCTGTATG 779
|||||

RESULT 2

AX254678 LOCUS AX254678 779 bp DNA linear PAT 10-OCT-2001

DEFINITION Sequence 2 from Patent WO0170247.

ACCESSION AX254678

VERSION AX254678.1 GI:16074345

KEYWORDS .

SOURCE Salmonella typhimurium.

ORGANISM Salmonella typhimurium

REFERENCE 1 (bases 1 to 779)

AUTHORS Lowery,D.E. and Kennedy,M.J.

TITLE Salmonella vaccine materials and methods

KEYWORDS pathogenicity island; secretion system apparatus; ssaQ gene; ssaR gene; ssaS gene; ssaT gene; ssaU gene.

SOURCE Salmonella typhimurium.

ORGANISM Salmonella typhimurium

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.

REFERENCE 1 (bases 1 to 5484)

AUTHORS Hensel, M.

TITLE Direct Submission

JOURNAL Submitted (09-AUG-1996) M. Hensel, Max von Pettenkofer-Institut, Lehrstuhl fuer Bakteriologie, Pettenkofer Strasse 9a, 80336 Munich, FRG

REFERENCE 2 (bases 1 to 5484)

AUTHORS Hensel, M., Shea, J.E., Baumlner, A.J., Gleeson, C., Blattner, F. and Holden, D.W.

TITLE Analysis of the boundaries of Salmonella pathogenicity island 2 and the corresponding chromosomal region of Escherichia coli K-12

JOURNAL J. Bacteriol. 179 (4), 1105-1111 (1997)

MEDLINE 97175536

FEATURES

source Location/Qualifiers

1. .5484

/organism="Salmonella typhimurium"

/strain="LT2"

/db_xref="taxon:602"

1. .653

/gene="ssaQ"

<1. .653

/gene="ssaQ"

/codon_start=3

/transl_table=11

/product="secretion system apparatus, SsaQ"

/protein_id="CAA68198.1"

/db_xref="GI:1526984"

/db_xref="SWISS-PROT:P74860"

/translation="CSNLPQLALHKKWVEEHFHSIIFTWPTGFLRNIVGELSAR QQIVPAPPVVPVYSGWCOLTLIELEIRIGMGVRIHCFDRIQLGFAIOLPGGIYAR VLLLEDNWKFDVLDIETLLASGSPMSKSDGTSVELEQIQOVLFEVGRASLEIG QLRUKTGVDYLPVGGCFAPVTVIRVNDRIIGOGELIACGNFVRITRWYLCNKTA" 721. .1368

/gene="ssaR"

721. .1368

/gene="ssaR"

/codon_start=1

/transl_table=11

/product="secretion system apparatus, SsaR"

/protein_id="CAA68199.1"

/db_xref="GI:1526985"

/db_xref="SWISS-PROT:P74890"

/translation="MSLPDPSPLQILGILFLLSLPLIIVMGTSFLKLVVFSILRNAL GIQQVPPNIALYGLALVLSLFINGPTLLAVKRRHPVQVAGAPFTWSEWDSKALAPYR QFLQKNSEKEANYFRNLKRTWPEDIKRIKPDLSLLIIPAFVTSQLTQAFRIGLLI YLPFLAIDLILISNILLAMGMVMVSPMTISLPFKLLIFLLAGGWDLTJLAQLVQSFS" 1365. .1631

/gene="ssaS"

1365. .1631

/gene="ssaS"

/codon_start=1

/transl_table=11

/product="secretion system apparatus, SsaS"

/protein_id="CAA68200.1"

/db_xref="GI:1526986"

/db_xref="SWISS-PROT:P74891"

/translation="MNDELTFQVTLQLLWLVFTSMFVPLVASVVGIVSLVQALTOI QDQTLQFMIKLLATAITLMSVPLWSGILLNTRQIMLRIGEHG" 1632. .2411

/gene="ssaT"

1632. .2411

/gene="ssaT"

/codon_start=1

/transl_table=11

/product="secretion system apparatus, SsaT"

/protein_id="CAA68201.1"

/db_xref="GI:1695195"

/db_xref="SWISS-PROT:P96068"

/translation="MAQQVNEMLIALAVAFIRPLSLSLLLPLLKSGSILGAALLRNGVL MSLTFPILPIYQOKIMMHIGKDYSLGLVTEVIGFSIGCAAVPFWVDNMGFLL DTLRGATMGTFINTEAETSLFGLFQSLFVIFCMVLMADLALGLLRSNOQLNVFFFS RTLFDDQFLKYIOAEWRTLYQICISFSLPAICMVLADLALGLLRSNOQLNVFFFS MPLKSILVLLTLTISFPYALHHYLVESDKFYIYIKDWPPSV" 2408. .3466

/gene="ssaU"

2408. .3466

/gene="ssaU"

/codon_start=1

/transl_table=11

/product="secretion system apparatus, SsaU"

/protein_id="CAA68202.1"

/db_xref="GI:1695196"

/db_xref="SWISS-PROT:P96069"

/translation="MSEKTEQPTKKLRDGRKEGVVKSIEITSLFOLIALYLIPIHF TEKMLILIESITFTLQVKNPFSYALQLSHALIESLTSALIFGAGVIVATVGVVF LOGVVIAASKAIGFKSEHNPVSNPKQIFSLHVVLELCKSLKVIMLSLIFAPFVYV ASTFRALPYCGLACGLVYSSSLIKHLWGVWVFYIVGILDSFOYTKIRDLKMSKD DVKQEHKDLLEGDPQMKTRRREMOSEIQSGSLAOSVKQSVAVVRNPTTHIACVGHPTD MPIPRVLEKGSDAOANYIVNIAERNICIPVENVELARSLFFEVERGDKIPETLFEPA ALLRMVMKIDYAHSTETP" 1496 a 1176 c 1277 g 1535 t

BASE COUNT 1496 a 1176 c 1277 g 1535 t

ORIGIN

Query Match 97.9%; Score 763; DB 1; Length 5484;

Best Local Similarity 98.7%; Pred. No. 7.2e-174;

Matches 769; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 atgggcacacaggtgaaatgagctgcttattgctgctgctgctgctgcttatttcgaccattc 60

Db 1632 ATGGCACACAGGTAAGTGGCTTATTCATTTGGCTGGCTTTTATTCGACCATG 1691

Qy 61 agccttcttatttatttccctatttataaaagtgagcagtttagggcgctctctttagct 120

Db 1692 AGCCTTCTCTTATTACTTCCCTTATTTAAAGTGCAGTTTAGGGCCGACCTTTTACGT 1751

Qy 121 aatggcgctgctatgtcaacttacccttcccatattaccatcattaccagcagaagatt 180

Db 1752 AATGGCGTCTTATGTCTACCTTACCTTCCGATATTTACCAATCATTTACCAGCACAAGATT 1811

Qy 181 atgagtcataattggttaagattacagttggttagggttagtcacgggagaggtgattatt 240

Db 1812 ATGATGCATATTGGTAAAGATTACAGTTGTTAGGTTAGCTACTGGAGAGGTGATTATT 1871

Qy 241 ggttttttaattgggttttggcgcggttcccttttggcggttgcgtatgatatggcggttt 300

Db 1872 GGTTTTCAATTGGGTTTGTGGCGCGTTCCTTTTGGCCCGTTGATATGGCCGGGTTT 1931

Qy 301 ctgcttgatctctttagctggtggtggtggtggtggtggtggtggtggtggtggtggtggt 360

Db 1932 CTGCTTGATACATTTACGTGGCGCGCAATGGGTACCATATTCAAATCTACATAGAAGCT 1991

Qy 361 gaaacctcaacttttggctgcttcttcagccagtttttggctggttattttcttcttataagc 420

Db 1992 GAAACCTCACATTTTGGCTTGTCTTTCAGCCAGTCTCTTGTGTGTTATTCTTTTATNAGC 2051

Qy 421 ggcggtcatggagtttattataaacattctgtatgagtcattatcatttaccaccaggg 480

Db 2052 GGCGGCATGGAGTTTATATAACATCTCTGTATGAGTCATATCAATATTTACCACAGGG 2111

Qy 481 cgtactttattatttgaccgggcaatttttataataatattccaggcagagtgagagacgct 540

Db 2112 CGTACTTTATTATTGACACCAATTTTAAATAATATATCCAGGCAGAGTGGAGAACGCTT 2171

Qy 541 tatcaattatgtgtcagtttctctctctccctccataataatgtatgtattagcgatctcg 600

Db 2172 TATCAATTATGATACAGCTTCTCTCTCTCTCCATAAATATGTATGTATTAGCGATCTG 2231

Qy 601 gcttttaggtcttttaaatcggtcggcacacaacaatgaaatgtgtttttcttctcaatgcgg 660

VERSION	A51671.1	GI:2304475
KEYWORDS	Salmonella typhimurium.	
SOURCE	Salmonella typhimurium	
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.	
REFERENCE	1 (bases 1 to 297)	
AUTHORS	Holden,D.W.	
TITLE	IDENTIFICATION OF GENES	
JOURNAL	Patent: WO 9617951-A 15 13-JUN-1996;	
	RPMS TECHNOLOGY LTD (GB)	
COMMENT	Other publication AU 4121996 960626.	
FEATURES	Location/Qualifiers	
source	1..297	
	/organism="Salmonella typhimurium"	
	/db_xref="taxon:602"	
BASE COUNT	99 a 67 c 55 g 76 t	
ORIGIN		
Query Match	26.6%; Score 207.4; DB 6; Length 297;	
Best Local Similarity	97.2%; Pred. No. 8.2e-40;	
Matches 211; Conservative	0; Mismatches 6; Indels 0; Gaps 0;	
Qy	1 atggcacacaggtaaatgagtggtctattgcattggctggtgctgtttatttcgaccatta 60	
Db	224 ATGGCACAACAGGTAATGAGTGGCTTATTGCATTGGCTGGCTTTATTTCGACCATTG 165	
Qy	61 agcctttctttattactccctattataaaagtggcagtttagggcgccgtcttttacgt 120	
Db	164 AGCCTTTCTTTATTACTCCCTATTAAAAAGTGGCAGTTTAGGGCCGCACATTTTACGT 105	
Qy	121 aatggcgcttatgtcacttaccctttcccataattaccacattattaccagcagaagatt 180	
Db	104 AATGGCGTGTATGTCACCTACCTTCCGATATTACCAATCATTTACCAGCAGAAGATT 45	
Qy	181 atgatgcataattggtaaaagattacagtttgtagggt 217	
Db	44 ATGATGCATATTGGTAAAGATTACAGTTGGGTAGGTT 8	
RESULT	8	
AX001605/c		
LOCUS	AX001605	297 bp DNA linear PAT 10-MAR-2000
DEFINITION	Sequence 15 from Patent EP0889120.	
ACCESSION	AX001605	
VERSION	AX001605.1	GI:7241734
KEYWORDS	Salmonella typhimurium.	
SOURCE	Salmonella typhimurium	
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.	
REFERENCE	1 (bases 1 to 297)	
AUTHORS	Holden,D.W.	
TITLE	A micro-organism having reduced adaption to a particular environment	
JOURNAL	Patent: EP 0889120-A 15 07-JAN-1999;	
	IMP COLLEGE INNOVATIONS LTD (GB)	
FEATURES	Location/Qualifiers	
source	1..297	
	/organism="Salmonella typhimurium"	
	/db_xref="taxon:602"	
BASE COUNT	99 a 67 c 55 g 76 t	
ORIGIN		
Query Match	26.6%; Score 207.4; DB 6; Length 297;	
Best Local Similarity	97.2%; Pred. No. 8.2e-40;	
Matches 211; Conservative	0; Mismatches 6; Indels 0; Gaps 0;	
Qy	1 atggcacacaggtaaatgagtggtctattgcattggctggtgctgtttatttcgaccatta 60	
Db	224 ATGGCACAACAGGTAATGAGTGGCTTATTGCATTGGCTGGCTTTATTTCGACCATTG 165	

```

gene
complement(804.. .1265)
/gene="asnC"
/note="YPO0002"
CDS
complement(804.. .1265)
/gene="asnC"
/note="Similar to the Escherichia coli regulator of Ansa:
AsnC SW:ASNC_ECOLI (P03809) (152 aa) fasta scores: E(): 0,
84.8% id in 151 aa"
/codon_start=1
/transl_table=11
/product="regulatory protein"
/protein_id="CAC88869.1"
/db_xref="GI:15978117"
/translation="MSEYQIDNLDRLSILKALMENARTPYAELAKNLAVSPGTIHVRV
EKMRAQGITACVHVNPQLGVDVCCFIIILKSADYPSALKKLESLEVEVEAYIT
TGHYSIFIRVMKSIDALQQVLINKIQTIDEIQSTETLISLQNPIMRTIYP"
complement(870.. .1175)
/gene="asnC"
/note="Pfam match to entry PF01037 ASNC_trans_reg, AsnC
family, score 63.50, E-value 4.4e-15"
complement(1113.. .1193)
/gene="asnC"
/note="PS00519 Bacterial regulatory proteins, asnC family
signature."
complement(1131.. .1196)
/gene="asnC"
/note="Predicted helix-turn-helix motif with score 1363
(+3.83 SD) at aa 6-27, sequence TPAELAKNLAVSPGTIHVRVE"
1435.. .2427
/gene="asnA"
/note="YPO0003"
1435.. .2427
/gene="asnA"
/EC_number="6.3.1.1"
/note="Similar to Escherichia coli aspartate-ammonia
ligase AsnA SW:ASNA_ECOLI (P00963) (330 aa) fasta scores:
E(): 0, 78.5% id in 330 aa"
/codon_start=1
/transl_table=11
/product="aspartate-ammonia ligase"
/protein_id="CAC88870.1"
/db_xref="GI:15978118"
/translation="MKKQIQKQQQISFVKSSFRQLEQQGLIEVQAPILSRVGCN"
QDNLSSGKAVQKVSPLDSTFEVVSIAKWKRTLGRFDGADQGVYTHMKALRPD
EDRLSATHSVYDQWDERVNGDGERNLAYLKSTVKNYAAIKETEAAISAEFGKPF
LPDHQITIHSESLRARPDLDAKGRERAIKELGAVFLIGIGKLADQSHDVRAPDY
DWTSPSAEGFSLNGDIIVNPNILEDAFEISSMGIRVDAEALKROLALTGDEDRLEL
EWHQSLRGMPTIGGGIGQSRVYMLLLQKHQIGQVQCGVWGPEISEKVDGLL"
complement(2526.. .3992)
/gene="YPO0004"
complement(2526.. .3992)
/gene="YPO0004"
/note="Similar to Vibrio cholerae hypothetical protein
VCA0762 TR:Q9KLI2 (EMBL:AE004404) (481 aa) fasta scores:
E(): 0, 34.1% id in 496 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC88871.1"
/db_xref="GI:15978119"
/translation="MLSLATDMLLISSEGLIEEMVVGLLAAPOLATFEKFPRIKR
ALMKDIPGWKQNLQKRIREASVPGLANEFLYQQSLLEDSPQFYAHLPDIVAQLQDL
HSPFATQAKLVQTADLAKNPPGDSLQTLFLQRWRVSLIQTITIHQLLEQEREQL
LAELQRLALSGALEPILTDNNAAGRLWMSQGHLSQDYQLLLQYDGLQQQPELI
RLAEQGRSRSAKAPADARYEPTVMVRQDPSVEVSGIHQSNIDILRLPTELVM
LGMSELEFEFVRLLERLITYRQGDNNQKQTPQRPVSLKONDEQPRGPTVCVDTIS
CSMGCGNEQCAKAFCLALRLIALADNRRCYIMLFATEIHYELSDANGIEQAIRFLNQ
HFRGGTDLAACALANTLNKMDREWDADAVIISDFIAQLRPELVKRIKIQQAQHQHR
FHAVASVYKPGIMRIEDHWRDTSLSRLIRWRK"
complement(3996.. .5549)
/gene="YPO0005"
complement(3996.. .5549)
/gene="YPO0005"

gene
complement(804.. .1265)
/note="Similar to Escherichia coli hypothetical protein
yIen SW:YIEN_ECOLI (P31473) (498 aa) fasta scores: E(): 0,
67.8% id in 497 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC88872.1"
/db_xref="GI:15978120"
/translation="MAOSSOLAERISRLSHALESGLYERQEAIRCLLLAALSGESVFL
LGPPGIAKSLIARRLKFAFRHARAFYELMTPEESTPEEVGPLSQALKEGRYQRMGTG
GYLPEAEIVFLDEIKWAGPAINTLTAINERRFRNGDRSDSIPMLRLVTAQELPDA
DSSLEALYDLIRLWLDROVQEKNEFSLLISRONENHNPNVAENLITDEEFHQWQPL
IDKITLDPHCCELIFOLRQRLSALEHAPYVSDRRKKALRLQASAFFSGRDEITP ID
LILLDKLWHDLNSFKLLQQLEQLITEGQQYQQQQLMKQDIINSKWLQHQOQSDHQ
ALTVPVQSGMFRKQAYALPDLNLTSTLTLLQKPLNLDHDIQVNHLOVDKEALQWLN
KGGALRAKLVGYAQSIDAIEDDOLHIILDVSRQPSLTSLPGLCATTTVPPELLAL
TKLESTLAEQRRLFSQHQPCLFTPSSWLAKTEASLLQVVEQLQFQIQFIQFIQQRK
FOOKKHSGH"
5823.. .7691
/gene="kup"
/note="YPO0006; trkD"
5823.. .7691
/gene="trkD"
/note="Similar to Escherichia coli low affinity potassium
uptake system protein Kup SW:KUP_ECOLI (P30016) (622 aa)
fasta scores: E(): 0, 84.2% id in 622 aa"
/codon_start=1
/transl_table=11
/product="potassium transport protein"
/protein_id="CAC88873.1"
/db_xref="GI:15978121"
/translation="MSTEKQYLSAVTLAAIVGVYDICTSPLYTLRRCFSGHYGFVDV
RPDVVFGFLSLIFMLILVSVKLYTVMRADNAGEGGLTILMSLAGNNTSRATSLI
VVLGLIGSFYGEVYITPAISVMASIEGLEIAPALDPYIVPCSIATLTLFVIOKH
GTGSGVKLPAPVMLVWFLTLALGLRSIIANPEVLAALNPKWASIFVEYKSVSFFALH
GAVLAITGEALYADMGHFGKFPRLAWFTVLPVSLVLYFCQGALLKKPEAIKNP
FFLLAPDWALIPLLILATLATVIASQAVISGVFSLTRQAVRLGLYLPWRITHTSEMES
GOIYIPVINTLYLAVLVLIIGFERSNLAAYGIAVTGTMTSILFTVAVKNNHW
NRFLVFLMLLIIDIPMSANVILKLSGGWPLSLGLVMFTIMTWKSERFLRR
MHEHNSLEAMIASLEKSPVPVGTAVYMSRAMNVIIPALLHNLKINKVLHERVLL
TMRTDVPPVHVNERVTIEQLSPTFWRVARYGWRETPNVAEIFHRCGLELSCQMMH
TSFFMHSHELILTKRPWHLFLRGKLFIALSRNALRAPDQFEIPPNRVIELGTQVEI"
5841.. .5909
/gene="kup"
/note="One of 12 probable transmembrane helices predicted
for YPO0006 by TMHMM2.0"
5952.. .6020
/gene="kup"
/note="One of 12 probable transmembrane helices predicted
for YPO0006 by TMHMM2.0"
6126.. .6194
/gene="kup"
/note="One of 12 probable transmembrane helices predicted
for YPO0006 by TMHMM2.0"
6222.. .6290
/gene="kup"
/note="One of 12 probable transmembrane helices predicted
for YPO0006 by TMHMM2.0"
6324.. .6392
/gene="kup"
/note="One of 12 probable transmembrane helices predicted
for YPO0006 by TMHMM2.0"
6450.. .6518
/gene="kup"
/note="One of 12 probable transmembrane helices predicted
for YPO0006 by TMHMM2.0"
6555.. .6623
/gene="kup"
/note="One of 12 probable transmembrane helices predicted
for YPO0006 by TMHMM2.0"

Query Match 23.2%; Score 180.8; DB 1; Length 288050;
Best Local Similarity 54.5%; Pred. No. 1.3e-33;

```

Matches 384; Conservative 0; Mismatches 317; Indels 3; Gaps 1;

QY 23 ggcttattgcatggtggtgctttttattgacacattgaagcctttttttattacttcccc 82
 Db 271625 GTCTAACCGGCACTCGCATATGACCATGATGGCGCTACGGCTATCTTACTGATATACCGC 271684

QY 83 tattaaaaagtgccagcttttggggccgcctttttacgtaatggcgtgttatcactta 142
 Db 271685 TGTTTACCGCGCGCACTTTGGGGAGCAGCTTATTGCGTAATGGCTGATTGTTGCCATCG 271744

QY 143 cctttcccatattaccaatcatttaccagcagaagattatgatgcataattggtaaagatt 202
 Db 271745 CGCTGCGGTCACGCGCTATTTTATCCGCACCATCATCACCATACTAGTACCCGGT-- 271802

QY 203 acagttggttagtgtagtcacggagagtgattgattggttttttttaattgggtttttgtg 262
 Db 271803 -CACCTGGATTGGGGTGTATGACGAGGACTACTCATTTGGGGTGGTGGTGTTCG 271861

QY 263 cggcggttcccttttggggcgttgatagcggggttttctgcttgatcattttacgtggcg 322
 Db 271862 CGCATGTCCTTTCGGGCTATGAATATGGCGGTTTCTCGATCGATACCTTTACGGGGGG 271921

QY 323 cgacaatgggtacgataattcattacataagaagctgaacacotcacttttttgggttgc 382
 Db 271922 CAACGATGTCCACGTTGTTTAAACCGCGTATGGCGTCAATCTCCCTTTTGGTGTGC 271981

QY 383 ttttcggcagtttttgggtgtgtatttttcttataagcggcgcatggagtttatattaa 442
 Db 271982 TGTTTACCCAAATATTAAACGGTACTGTTCTCTGATTCTGCGGTTTAAATCAGGTGTTGG 272041

QY 443 acattctgtatgagtcataatcatttaccacaggcgctactttatttattgaccgac 502
 Db 272042 CGCGGTGTACGGCTCGTACCATAGCTTACCCATCGCCAGGTTATCCAGCTGCTGCAG 272101

QY 503 aatttttataatataccagcgagtgaggagacgctttatcatcaattatgctcagtttct 562
 Db 272102 ATCTATTGCTCTTCTGCAAACTGAGTGGCAGATGATGTCGACCTTTCCTGTGCTTGG 272161

QY 563 ctcttcctgccataatgatgtatggtattagccgactctgcttttaggtcttttaaatcggt 622
 Db 272162 CCTACCGGCACTGTGGTCAATGGTGTGGCTGATCTCTTTTAGGATTAATTAACCGCT 272221

QY 623 cggcacacaattgaatgtgtttttcttctcaatggcgctcaaaagtatatattggttctac 682
 Db 272222 CGGCCCGCAACTGAATGTCTTTTCTGGGATGCCCATTAAGCGCACTGCGGCTTT 272281

QY 683 tgacgctcctgatctcattcccttattgctttcattcactatttg 726
 Db 272282 TTCTCTTACTGATCAGCTGCCTTATGGCCTGCATCAGTACCTG 272325

RESULT 10

AF022236

LOCUS AF022236 35624 bp DNA linear BCT 25-APR-1998

DEFINITION Escherichia coli strain E2348/69 pathogenicity island, rOrf1 (rOrf1, rOrf2 (rOrf2), EscR (escR), EscS (escS), EscT (escT), EscU (escU), CesD (cesD), EscC (escC), EscV (escV), SepZ (sepZ), EscV (escV), EscN (escN), SepQ (sepQ), Tir (tir), OrfU (orfU), intimin (eae), EscD (escD), SepL (sepL), EspA (espA), EspD (espD), EspB (espB), EscF (escF), EspF (espF) genes, complete cds.

ACCESSION AF022236

VERSION AF022236.1 GI:2897961

KEYWORDS Escherichia coli.

SOURCE Escherichia coli

ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 35624)

AUTHORS Elliott,S.J., Wainwright,L.A., McDaniel,T.K., Jarvis,K.G., Deng,Y.K., Lai,L.C., McNamara,B.P., Donnenberg,M.S. and Kaper,J.B.

TITLE The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic Escherichia coli E2348/69

JOURNAL Mol. Microbiol. 28 (1), 1-4 (1998)

98254123

2 (bases 1 to 35624)

REFERENCE Elliott,S.J., Wainwright,L.A., McDaniel,T.K., Jarvis,K.G., Deng,Y.K., Lai,L.C., McNamara,B.P., Donnenberg,M.S. and Kaper,J.B.

AUTHORS Direct Submission

TITLE Submitted (03-SEP-1997) Center for Vaccine Development, University of Maryland School of Medicine, 685 W Baltimore St, Baltimore, MD 21201, USA

JOURNAL

REFERENCE 3 (bases 1 to 35624)

AUTHORS Elliott,S.J. and Kaper,J.B.

TITLE Direct Submission

JOURNAL Submitted (19-FEB-1998) Center for Vaccine Development, University of Maryland School of Medicine, 685 W Baltimore St, Baltimore, MD 21201, USA

REMARK Sequence update by submitter

REFERENCE 4 (bases 29248 to 30303)

AUTHORS O'Connell,C. and Donnenberg,M.S.

TITLE Direct Submission

JOURNAL Submitted (05-MAR-1998) University of Maryland School of Medicine, 685 W Baltimore St, Baltimore, MD 21201, USA

REMARK update of Orf23 to SepL submitted by C. O'Connell and M.S. Donnenberg

REFERENCE 5 (bases 34121 to 34741)

AUTHORS McNamara,B.P. and Donnenberg,M.S.

TITLE Direct Submission

JOURNAL Submitted (05-MAR-1998) University of Maryland School of Medicine, 685 W Baltimore St, Baltimore, MD 21201, USA

REMARK update of Orf30 to EspF submitted by B.P. McNamara and M.S. Donnenberg

COMMENT On Feb 21, 1998 this sequence version replaced gi:2865268.

FEATURES

source

1. 35624

/organism="Escherichia coli"

/strain="E2348/69"

/serotype="O127:H7"

/db_xref="taxon:562"

/note="isolated from an outbreak in children in U.K. in 1969; pathogenicity in adults demonstrated in 1977"

/gene="rorf1"

/complement(676..1494)

/complement(676..1494)

/gene="rorf1"

/codon_start=1

/transl_table=11

/product="rOrf1"

/protein_id="AAC38362.1"

/db_xref="GI:2865269"

/complement(1622..2818)

/gene="rorf2"

/complement(1622..2818)

/gene="rorf2"

/codon_start=1

/transl_table=11

/product="rOrf2"

/protein_id="AAC38363.1"

/db_xref="GI:2865270"

/translation="MILVAKLFTNQIGSLMINGLNDSASLVDAAMKVNVSFKKS WDEMSAEKLLKVLFGLMNPYSRSEROSFOELLTVLEPVLPNEIGRVSARFSDG SSLRISVTNSESIEARIPTDNEKITVLLSEQNRLQLSLPIDRHMPYIQVHRAI.SE MDLTDTSMRNLIGFTSKLSTLIPHNAOTDPLSCPTPFSSIEMDTGCGLGNKLSLN GVDIPANAQMLLRDALGLKDTSHSSPSRVNIDHISRHDAEQIARESSGSDNKAQVVE FLCHPEAATAICSAFYQSFNVPAALTHTHERISKASEYNAERSLDTPNACINISISQS DGNIVYTSHTGVLINAPEDRPNEMGLTNRNRTSYEVPQGVKCTIDEMVRALQPRYAASE TYLQNT"

3575..3700

/note="IRU/ERIC (intergenic repeat unit/enteric repetitive intergenic consensus) sequence"

4070..4459

repeat_region

gene

CDS


```
/codon_start=1
/transl_table=11
/product="Orf1"
/protein_id="AAC38364.1"
/db_xref="GI:2865271"
/translation="MRRLFMMETNSHTTSPYQLIEQIAVLQAEAKRLREQEVQSV
IESIQKQIYYNITLQELGYTNVDDGLARRNSSRGVYRNEEGTWSGVGRQPRWLK
EALNGMKEDFLVKDTEETIPLKNI"
4474..4692
CDS
/codon_start=1
/transl_table=11
/product="Orf2"
/protein_id="AAC38365.1"
/db_xref="GI:2865272"
/translation="MITTELEDEIINKKEAANVFIEKINDKKNHEKMKHPLDKVT
YNEAKELIACDAAIRTEIMRIRINK"
4696..5019
CDS
/codon_start=1
/transl_table=11
/product="Orf3"
/protein_id="AAC38366.1"
/db_xref="GI:2865273"
/translation="MSIVSQTRNKELLDKKIRSEIEIAIKKIIABFDVVKESVNELSEK
AKTPOAEKLNKLIEGYTYGEERKLYDSALSKIEKLIETLSPARSKSOSTMNQNRN
NRKIV"
5016..5615
CDS
/codon_start=1
/transl_table=11
/product="Orf4"
/protein_id="AAC38367.1"
/db_xref="GI:2865274"
/translation="MTIFNKIDYNYKLIINYPIMNVHDEWLDGLTVGNQVSPRRLET
SSTRNQMLKEIQIDKISGVELSKIDKEGFLVQSIGKIRLLALSSALFDIOCPDYI
FSRLYRTELIRIGYQNVKLSFYWGQGGCKPEYGEERCAELIKYGAGNLEWLFADN
PLWTIVYLLPSGEIKPHTINDLFLNRLNKILLPYETL"
5560..6255
CDS
/codon_start=1
/transl_table=11
/product="Orf5"
/protein_id="AAC38368.1"
/db_xref="GI:2865275"
/translation="MIYFLDLIKFYCLMKLYEFKNIEIDLVLTEIIPEDKIQEIIQ
SDIILKARKTYEHLRARRKSKELKIESRKKIARKMIAMRIRIRKNNKIKLDKEVN
QSIKWQDIOAIELVLMQDIMNKVHLSLTNALHSLDTSRINWDDLLNEVRETLSNN
NIVGAIKITKPNPDIKLDPGEANNIQLINDANTPHNKIILIENEYIRITLDPLEQISILL
NSFKDNYLSIIQE"
6260..6913
gene
/genes="escr"
6260..6913
CDS
/codon_start=1
/transl_table=11
/product="Escr"
/protein_id="AAC38369.1"
/db_xref="GI:2865276"
/translation="MSQLMTIGSOPIFLIIVFFLLSLPIFVVIGTSFLKISIVLGIL
KNALGQOVPPNMAITSVLITMFIIQINDNISOEPIINYTSDSFQKVDERKI
LSPYRGILKEKNDNVEFFERAQKILGNSETILKDSILFILLPAFTMGQLEAAFKIG
FLLYLPFIADLIISNILLALGMMVSPVTISIPFKILLFLVGGWQKLFEPFLVYN"
6913..7182
gene
/genes="escs"
6913..7182
CDS
/codon_start=1
/transl_table=11
/product="Escs"
/protein_id="AAC38370.1"
/db_xref="GI:2865277"
/translation="MDTGFVQLCVQTFWIIIFILSLPTVIAASVIGIISIVQAITQL
QDQLPFLKIIIAVFATLALTYHMMGTIIINFSIIIFEMIPKVG"
7182..7958
gene
/genes="escr"
7182..7958
CDS
```

```
/gene="escr"
/codon_start=1
/transl_table=11
/product="Escr"
/protein_id="AAC38371.1"
/db_xref="GI:2865278"
/translation="MNEIMTVIVSSYCIURLPLGMFIILPIFSTGVLNSNFIRNSIMI
AFTLPIIYNTFSEKLPSTGQFQLTGIVAKISIGFFIGLISFTILFWADAAGQIIDT
LRGSTIISIFNPSISDSSITGVILYQFISVIFVHGGIQSILDKLXLYSLEILLOAD
IAFNRALIDELFSLWDSFIKMLSESVPMIIGIFLDCMGFGFLNKTAPQLNFTLSLP
VKSIIAIFILLLVHVPDFITANIHSIIDRSLPFSMINE"
7951..8988
gene
/genes="escu"
7951..8988
CDS
/codon_start=1
/transl_table=11
/product="Escu"
/protein_id="AAC38372.1"
/db_xref="GI:2865279"
/translation="MSEKTEKPTPKKLRDLKKKGDVTKSEVMAAVQSILFSFFSLY
GMSFFVDIVGLVNTTIDSLNRPFYAIRIELGAVLNIFLLYILPISLIVFVGVTGTY
SQIGFFAVEKIKPSAQKISVKNLKNIFSVKSFELLKSVFKLVIIVLIFPMGHYSY
ANEFANFTGLNAYQALVYVAFVFLWKGVLFGYLLFSVDFWFKHGLKKMKMSKD
EVKREAKTDGNPEIKGERRLHSEIQSGSLANNIKKSTVIKNPTHTAICLYVKLGE
TPLPVIETGDKAKALQIIKLAELYDIPVIEDIPLARSLYKNIHKGYITEDFFEPVA
QLIRIAlDLdy"
complement(8985..9443)
CDS
/codon_start=1
/transl_table=11
/product="rOrf3"
/protein_id="AAC38373.1"
/db_xref="GI:2865280"
/translation="MKKIILISNVCNASDCFEITGKAVNIDPILKAIWNBSK
```

```
Query Match      15.4%  Score 119.6;  DB 1;  Length 35624;
Best Local Similarity 52.2%  Pred. No. 8.4e-19;
Matches 266;  Conservative 0;  Mismatches 244;  Indels 0;  Gaps 0;

QY  208  tgggttaggggttagtcaccggagaggtgattattggttttttaattgggtttgtgcgcgcg 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  7383  TTGACAGAAATAGTGTCTAAAGAATAAGTATTGGTTTTTTCATTTGGGTATCATATTACT 7442
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  268  gtcccttttgggcggttgatagcggggtttctgctgtgatactttacgtggcgcgaca 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  7443  ATTCTTTTTCGGCAATAGATGCGGCTGGCGAGATTATTGATACCTAAGAGGTTCAACA 7502
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  328  atgggtacgatattcaattctacaatagaaagcgaacctcaactttttgggttgcctttc 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  7503  ATATCTTCAATTTTAAACCGTCCATAGTGATTTCATCTCTATCACTGGCGTTATTGTG 7562
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  388  agccagtttttgggtgattttcttataagcggcggaatgaggtttattataaacatt 447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  7563  TACCAATTATCTCTGTGATCTTTGTTATTATTCATGGGATACAAAGCATTCGGATAAG 7622
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  448  ctgtagagtcataatcaattttaccacaggcggtactttattatttaccggcaattt 507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  7623  CTATAATTATCCTACGAGATATTACCATTACAAACCCGATATTCGATTCAAATCGTGCCTTA 7682
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  508  ttaaaatatatccaggcagagtgaggaacgctttatcaattatgtgcagtttctctctt 567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  7683  ATAGATTTTTTGTGTTTCTCTATGGGATTCATTTATTAACTGATGTATCATTTTCAGTT 7742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  568  cctgccaataatgtatggttagcgcgattcgttctttaggtctttcaaatcggtcgcca 627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  7743  CCCATGATTATCGGTATATCTTATGTGATATGCGGTTTGGGTTTCTTAACAAAAACGCG 7802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  628  caacaattgaatgtgttttctctctcaatgcccgctcaaaagatatattggttctactgacg 687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  7803  CCTCAACTAAACGGATTATTCACATTATCACTGCCAGTAAAAAGTTTGATAGCAATCTTTATA 7862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  688  ctccatgatctcatccctcttctgctcttcat 717
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


Db	7863	TTGCTATTAGTATTATTCATGTTTTCCTGAT	7892	
RESULT 11	AE005597/c			
LOCUS	AE005597	11760 bp	DNA	linear BCT 21-MAR-2001
DEFINITION	Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 216 of 290.			
ACCESSION	AE005597	AE005174		
VERSION	AE005597.1	GI:12518472		
KEYWORDS				
SOURCE	Escherichia coli O157:H7 EDL933.			
ORGANISM	Escherichia coli O157:H7 EDL933			
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.			
AUTHORS	1 (bases 1 to 11760) Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.			
TITLE	Genome sequence of enterohaemorrhagic Escherichia coli O157:H7			
JOURNAL	Nature 409 (6819), 529-533 (2001)			
MEDLINE	21074935			
PUBMED	11206551			
REFERENCE	2 (bases 1 to 11760) Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.			
AUTHORS	Direct Submission Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA			
TITLE	Location/Qualifiers			
JOURNAL	1. 11760			
FEATURES	source			
	/organism="Escherichia coli O157:H7 EDL933"			
	/strain="EDL933"			
	/serotype="O157:H7"			
	/db_xref="taxon:155864"			
	/note="enterohaemorrhagic"			
	<1. .9623			
	/note="O-island #148; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"			
	129. .587			
	/gene="25131"			
	129. .587			
	/gene="25131"			
	/function="orf; Unknown function"			
	/note="Residues 1 to 152 of 152 are 100.00 pct identical to residues 1 to 152 of 152 from GenPept 118 :			
	gi 3414913 gb AAC31524.1 (AF071034) L0045 [Escherichia coli]"			
	/codon_start=1			
	/transl_table=11			
	/protein_id="AAG58843.1"			
	/db_xref="GI:12518473"			
	/translation="MKKIIISIIISNICYASDCFEITGKAYNIDPILKATAMNESK CKSGIKSNKNGTYDIGIMNSHLLSKFNISEDLLNDACINISVAGYILASN IKSRGNTWDVAGVAYNAGYENTPNAVELRRQYANKIYKYNKLKNEQIID"			
	complement(584. .1621)			
	/gene="escu"			
	/note="25132"			
	complement(584. .1621)			
	/gene="escu"			
	/function="transport; Transport of large molecules: protein, peptide secretion"			
	/note="Residues 1 to 345 of 345 are 100.00 pct identical to residues 1 to 345 of 345 from GenPept 118 :			
	gi 3414914 gb AAC31525.1 (AF071034) L0046 [Escherichia coli]"			
	/codon_start=1			
	/transl_table=11			
	/protein_id="AAG58844.1"			
	/db_xref="GI:12518474"			
	/translation="MSEKTEKPTPKRLDLKKKGDTVKSEYMAAVOSLILFSFFSLY GMSFVDIVGLVNTIDSUNRPFLYAIREILGAVLNIFLLYILPISLIVFGVTITGV SQIGFIVAEKIKPSAQKISVKNLKNIFSVKSGFELKSVFKLVIITVLIFYFMGHISY ANEPANFTGLNAYOALVVVAFVFLWKGVLFGVLLFSVDFDFWFKHGLKKMKMSKD EVKREAKTDGNPEIKGERRLHSEIQSGSLANNIKKSTVIYKNPHTHAIICLYYKIGE TPLPLVIETGKDKARALQIIKLAELYDIPVIEDIPLARTLYKNIHKQYITEDFFEPVA QLIRIADLDY"			
	complement(1614. .2390)			
	/gene="escr"			
	/note="25133"			
	complement(1614. .2390)			
	/gene="escr"			
	/function="transport; Transport of large molecules: protein, peptide secretion"			
	/note="Residues 1 to 258 of 258 are 100.00 pct identical to residues 1 to 258 of 258 from GenPept 118 :			
	gi 3414915 gb AAC31526.1 (AF071034) L0047 [Escherichia coli]"			
	/codon_start=1			
	/transl_table=11			
	/product="escr"			
	/protein_id="AAG58845.1"			
	/db_xref="GI:12518475"			
	/translation="MNEIMTVIVSSFCYILRPLGMFIILPIFSTGVLLSNFIRNSIM1 AFTPLIIVNTFSKLSGIFQLTGIALKEISGFFIGLSFTILFWAIDAAQIIDT LRGTSSIFNPSISDSSITVLYQFISVIFVHGGIQSILDKLYSIELPLQAD IAFNRALIDFLSLWDSFKLMLSFSPVMIIGIFLDCMGGFLNKTAPQLNVTLSLP VKSLIAFILLIVHVPFPIITANIHSIDIIRSLPSMINE"			
	complement(2390. .2659)			
	/gene="escs"			
	/note="25134"			
	complement(2390. .2659)			
	/gene="escs"			
	/function="transport; Transport of large molecules: protein, peptide secretion"			
	/note="Residues 1 to 89 of 89 are 100.00 pct identical to residues 1 to 89 of 89 from GenPept 118 :			
	gi 2865277 gb AAC38370.1 (AF022236) EscS [Escherichia coli]"			
	/codon_start=1			
	/transl_table=11			
	/product="escs"			
	/protein_id="AAG58846.1"			
	/db_xref="GI:12518476"			
	/translation="MDTGTVQLCVQTFWIIFILSLPTVIAASVIGIISLVQAITOL QDQTLFPLKIIIVPATLALTYHWMGTIINFSSIPEMIPKVNQ"			
	complement(2659. .3312)			
	/gene="escr"			
	/note="25135"			
	complement(2659. .3312)			
	/gene="escr"			
	/function="transport; Transport of large molecules: protein, peptide secretion"			
	/note="Residues 1 to 217 of 217 are 100.00 pct identical to residues 1 to 217 of 217 from GenPept 118 :			
	gi 3414917 gb AAC31528.1 (AF071034) L0049 [Escherichia coli]"			
	/codon_start=1			
	/transl_table=11			
	/product="escr"			
	/protein_id="AAG58847.1"			
	/db_xref="GI:12518477"			
	/translation="MSQMTIGSQPIELIIVFLLSLLPFVWIGTSPLKISIVLGIIL KNALGIQVPPNMALTSVLITMPSPIILQINDNISQEPINTDSDFPFQKVDKIL LSPYRGFLKTEKDNVEFFPRAAQKIKGNETILKDSLFILPAPFMQQLPAAPFKIG FLLYLPFIADLIISNILLALGMWVSPVTISIPFKILLFILVCGGWQKLPEFLLVN" complement(3317. .3931)			
	/gene="25136"			

CDS	complement(3317..3931) /gene="z5136" /function="orf; Unknown function" /note="Residues 1 to 204 are 99.50 pct identical to residues 28 to 231 from GenPept 118 : g1 3414918 gb AAC31529.1 (AF071034) L0050 [Escherichia coli]"	Db 2069 ATATCTTCAATTTTAAACCCGTCATAGTGATTCATCTTCTATCACTGGCGTTATCTTG 2010
		Qy 388 agccagtttttggtgtgtatttttttataagcgggcgatggagttttataataacatt 447
		Db 2009 TACCAATTTATCTCTGTGATCTTTGTTATTTCATGGTGGATACAAAGCATCCTAGATAAG 1950
		Qy 448 ctgtatgagtcataatatttaccaccagggcgctacttatttatttgaccggcaattt 507
		Db 1949 CTATATTTATCTACGAGATATTACCATTAACAGCGATATTGCATTCATTCGTCGCTTTA 1890
		Qy 508 ttaaatatataccagcgagagtgaggagacgctttatcaattatgtcagttctctctt 567
		Db 1889 ATAGATTTTGTGTTTCTCTATGGGATTCATTATTAAACGATGTTATCATCTTTTCAGTT 1830
		Qy 568 cctgccataatgtatggtattagccgacgtctggtcttaggtcttttaaatcggtcgga 627
		Db 1829 CCCATGATATATCGGTATATCTTATGTCATATGGCGTTTGGGTTCTTAAACAAACAGCA 1770
		Qy 628 caacaattgaatgtgtttttcttctcaatgcgcgtcaaaagtatatgttcttactgacg 687
		Db 1769 CCTCAGCTAAACGTATTTCACATTTATCACTGCGAGTAAAAAGTTTGATAGCAATCTTTATA 1710
		Qy 688 ctctgctctcattcccttattgctcttcatt 717
		Db 1709 TTGCTATTAGTTATTATTCATGTTTTCCTGAT 1680
RESULT 12		
AF071034/c		
LOCUS		
DEFINITION	AF071034 45325 bp DNA linear BCT 13-AUG-1998	
	Escherichia coli L0001 (yicJ) gene, partial cds: trna-Sec gene, complete sequence; CP4-like integrase (int), L0004, L0005, L0006, L0007, L0008, L0009, L0010, L0011, L0012, L0013, L0014, L0015, L0016, L0017, L0018 (escF), L0019, L0020 (espB), L0021 (espD), L0022 (espA), L0023 (sepl), L0024 (escD), L0025 (esae), L0026 (orfU), L0027 (tir), L0028, L0029, L0030, L0031 (sepQ), L0032, L0033, L0034 (escN), L0035 (escV), L0036, L0037 (sepZ), L0038, L0039 (escJ), L0040, L0041 (escC), L0042 (cesD), L0043, L0044, L0045, L0046 (escU), L0047 (escT), L0048 (escS), L0049 (escR), L0050, L0051, L0052, L0053, L0054, L0055, and L0056 genes, complete cds; and L0057 (yicL) gene, partial cds.	
ACCESSION	AF071034	
VERSION	AF071034.1 GI:3414870	
KEYWORDS	Escherichia coli.	
SOURCE	Escherichia coli	
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.	
REFERENCE	1 (bases 1 to 45325)	
AUTHORS	Perna,N.T., Mayhew,G.F., Posfai,G., Elliott,S., Donnenberg,M.S., Kaper,J.B. and Blattner,F.R.	
TITLE	Molecular evolution of a pathogenicity island from enterohemorrhagic escherichia coli O157:H7	
JOURNAL	Infect. Immun. 66 (8), 3810-3817 (1998)	
MEDLINE	98339885	
REFERENCE	2 (bases 1 to 45325)	
AUTHORS	Perna,N.T., Mayhew,G.F. and Blattner,F.R.	
TITLE	Direct Submission	
JOURNAL	Submitted (08-JUN-1998) Genetics Laboratory, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA	
FEATURES	Location/Qualifiers	
Source	1. 45325	
	/organism="Escherichia coli"	
	/strain="EDL93"	
	/strain="ATCC43895"	
	/serotype="O157:H7"	
	/db_xref="ATCC:43895"	
	/db_xref="taxon:562"	
	/note="Isolated from contaminated hamburger implicated in an outbreak of haemorrhagic colitis"	
	complement(<1..557)	
	/gene="yicJ"	
gene		

```
CDS      complement(<1. .557)
         /gene="yicJ"
         /codon_start=1
         /transl_table=11
         /product="L0001"
         /protein_id="AAC31536.1"
         /db_xref="GI:3414925"
         /translation="MVYRMYSQHPYSIKERQMKSEVLSVKEKIGYGMGGAASHIIF
         DNVMYMFYTFDNFGIPAGVGTMFVLVARALDAISDPCMGLLADRTSRWGKFRPWV
         LCGALPGIVCVLAYSTPDLNMCKMIYAAITYTLTLTYTVNIPYCALGCVITNDP
         TQRIQLQSRFVLATAGGMLSTVLDP"
tRNA      723. .902
         /note="L0002"
         /product="tRNA-OTHER"
         1083. .2264
         /gene="int"
         1083. .2264
         /gene="int"
         /note="L0003"
         /codon_start=1
         /transl_table=11
         /product="CP4-like integrase"
         /protein_id="AAC31482.1"
         /db_xref="GI:3414871"
         /translation="MALTDVVKTKAKPRPYKLDGGGMYLLINANGSKYWRMYRF
         AGKEKLSIGVYPDVTLADAKRSEARKILAAAGDGPCEAKKEKIALQMSLKNTPEA
         VAREHQTADVRSRLYRDEIIDTFEKDIFPYIGKRPIAEIKPMELLEALKRMEKRG
         LEKMKVRORCGEVFRYAIVTGRADYNAPDIASALATPKKVHFFLTANELPHFLTD
         LAGTGTIITKATQIIIMLTQVTRQELRFHAWEDIDPEAKLWEIPEAVMKMKRPHIVP
         PSEQVIALFKOLEPISKHPLVFIGRNDPRKPISKESINQVIELLYGKGRLTGHGFRH
         TMSITLHEQGFNSAWIEMQLAHVDMKNSIRGTYNHAQYLDGRREMMQYADYIDSSEL
         A"
CDS      2363. .2713
         /codon_start=1
         /transl_table=11
         /product="L0004"
         /protein_id="AAC31483.1"
         /db_xref="GI:3414872"
         /translation="MISSPQHKTGDLNMKTKRTFTPEFRLECAQLIVDKGYSYROAS
         EAMNVGSITLESVWRQLRRERQGIAPSATPIPTDQQRIRELEKQVRLEEHNTILKA
         TTALLMSDSLNGSR"
CDS      2726. .2908
         /codon_start=1
         /transl_table=11
         /product="L0005"
         /protein_id="AAC31484.1"
         /db_xref="GI:3414873"
         /translation="MTATQLSASVLLWKYTAAVTSTGENDAIRLIRRESGCAAKYAGR
         GTKVGALGRARWLKC"
CDS      2926. .3579
         /codon_start=1
         /transl_table=11
         /product="L0006"
         /protein_id="AAC31485.1"
         /db_xref="GI:3414874"
         /translation="MSRYRAGRLMKYLLNSSCPGKHQYKNARQEHCTPLNLERQFA
         VPEPRVWCGDITYIWAGNRWCYLAVMDFARRVIGHSLSANADTALISSALRMAYE
         VRGPDYVMPHSDQGSQYTKYQQLLWRYRIKQSVSRGNCWDSNPFERFSKTE
         WVPTDGTGDKVARQOISSYILNYNSVRPHYNGGLTPEESERNRYHFCYCKTASIT"
CDS      3830. .4204
         /codon_start=1
         /transl_table=11
         /product="L0007"
         /protein_id="AAC31486.1"
         /db_xref="GI:3414875"
         /translation="MKTLPDTHVREASRCPSPITIWOTLLGRLLDHVGTLTNDTPFA
         DERVEQHEAGISCLDAVNFLEKYLVRTDQPGFSACTRSQLSINSIDILRARRATG
         LMTRDNRVYNNTILGKHPEAK"
CDS      4201. .4692
         /codon_start=1
         /transl_table=11
         /product="L0008"
         /protein_id="AAC31487.1"
         /db_xref="GI:3414876"
         /translation="MMKLALTLEADSVNVQALNMGRIIVVDIGCVNLAEELINNVCDNGY
         SLRVVDESDRTSADCTPPETALTGIRCSFAHITETDNALWYLSLSHOTSDFCSEWIHP
         TNGYLLRTDAWSPVLRKLRIGLSKTPRRLVVVTLIRRYGVSILHLDASAGCLPLGPT
         FNV"
CDS      4704. .4901
         /codon_start=1
         /transl_table=11
         /product="L0009"
         /protein_id="AAC31488.1"
         /db_xref="GI:3414877"
         /translation="MTSLTPEAALDILIAWLQDNIDSESGIIFDNDDEKTDSAALLPC
         IEQVREDVRLRQLQLQNR"
CDS      4986. .5327
         /codon_start=1
         /transl_table=11
         /product="L0010"
         /protein_id="AAC31489.1"
         /db_xref="GI:3414878"
         /translation="MSGIILSRPEVDTGHTDVICSTIRHVTVVRNALQOQTETLIRQ
         LAEISVLTAADIGKRTALDRAMKQDFHCCGWLMEKPEAMKATIRNLDCETWRDLQOR
         SGMLSILMDAQ"
CDS      complement(5324. .5554)
         /codon_start=1
         /transl_table=11
         /product="L0011"
         /protein_id="AAC31490.1"
         /db_xref="GI:3414879"
         /translation="MKSPSPSHOVVDNFTAEFTGGIGFVIPFGPLKTLITPRSNIS
         SLLWSCSLNFMIFASLISGILSYSSERYHVS"
CDS      5528. .5713
         /codon_start=1
         /transl_table=11
         /product="L0012"
         /protein_id="AAC31491.1"
         /db_xref="GI:3414880"
         /translation="MRDWRFRHLISGMQADRLADLERMLHLLSGKPLPNNRGNITIN
         LDDHLQSVQKERYEDK"
CDS      6083. .6484
         /codon_start=1
         /transl_table=11
         /product="L0013"
         /protein_id="AAC31492.1"
         /db_xref="GI:3414881"
         /translation="MGTKVSDMKQNVTPGRRKGCNYPPEFKQQLVAASCEPCISISK
         LALENGINALLFKWRQOWREGKILLPSESPQLLPVTLDAAAEQPESLAEDEPETLSI
         SCEVTFRHGTLRFNGNVSEKLLTLTIQELKR"
CDS      6481. .6828
         /codon_start=1
         /transl_table=11
         /product="L0014"
         /protein_id="AAC31493.1"
         /db_xref="GI:3414882"
         /translation="MIPLPSGTKIWLVAGITDMRNGFNGLAAKVQTKALKDDPMGSHVF
         IFRGRSGSVKLLMSTGDGLCTKRLERGRFAMPSARDGKVFQLTQALMLLEGIDW
         ROPKRLTSLTML"
CDS      6878. .8416
         /codon_start=1
         /transl_table=11
         /product="L0015"
         /protein_id="AAC31494.1"
         /db_xref="GI:3414883"
         /translation="MNDISSDDIFLLKORLAEQEALIHAEQKLSNREREIDHLQAQL
         DKLRMNFGRSEKVSRRIAQMEADLNRLQKESDTLTGRVYDPAVQPLRQPTKRPFF
         PESLPDEKRLLPAAAPCCPNCGGSLSYLGEDTAQLELMRSAPRVIRTVREKHACTOC
         DAIVQAPAPSRPIERGIAGPLLARVLTSKYAEHTPLYRQSEIYGRQGVELRRSLSG
         WVDACRLLSPLEEALHGVTMDGKLHADDTPQVLLPFGNKTKTGRMLWYVYRDDRNA
         GSALAPVWEAYSPDKGIHPQTHLACFSGLQADAYAGFNEIYRNGGITEAACWAHA
         RRKINDVHVRIPSALTEALEQIGQLYALADI RGMFAQRLEAQRQTKPKLLKSL
         WLRKMKTLRSRHSLEKAFAYALNQWPAITYYANDGWVEIDNNIAENALRAYSICHKN
         WLFPGSDHGGGERGALLYSLTGCKLNDVDPESYLRHLVLCVIADWPVNRVSELLPWRIA
         LPAE"
         complement(9281. .10027)
```

```
/codon_start=1
/transl_table=11
/product="L0016"
/protein_id="AAC31495.1"
/db_xref="GI:3414884"
/translation="MLNGISNAASTLGRQLVGIASRVSSAGCTGFSAPOAVRLTPVK
VHSPSPGSSNVARTINVSQVTSPTSPSRAPPPPTSGQASGRPLPPIAQAALKE
HLAAEKSQGPALGFKPARQAPPPPTSGQASGRPLPPIAQAALKEHLAAEYKSGP
EALGFKPARQAPPPPTSGQASGRPLPPIAQAALKEHLAAEYKSGP
PPPTGSLPLPPLAQAALKEHLAAEYKSGK"
complement(10112..10390)

CDS
Query Match      15.1%; Score 118; DB 1; Length 45325;
Best Local Similarity 52.0%; Pred. NO. 2e-18;
Matches 265; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

Qy 208 tgggttaggttagtcaccgagagtgattggttttttaattgggtttttgtgcggcg 267
Db 36886 TTGACAGGAATAGCGCTAAAGAAATAAGTATTGGTTTTCATTTGGTTATCTTACT 36827

Qy 268 gtcccttttggcggttgatagcggggtttctgttgatactttacgtggtgcgaca 327
Db 36826 ATTCTTTTGGGCAATAGATGGCGCTGGACAGATTATTGATCTCTAAGAGGCTCAACA 36767

Qy 328 atgggtacgatattcaattctacaatagaagctgaacctctttttgtggtttcttc 387
Db 36766 ATATCTTCAATTTTAAACCGCTCATAGTGATTCTTCTATCTACTGGCGTTATCTTG 36707

Qy 388 agccagttttgtgttattttcttataagcggtgaggttttataataaacatt 447
Db 36706 TACCAATTTATCTGTGATCTTTGTTATTTCATGGTGGGATACAAAGCACTAGATAAG 36647

Qy 448 ctgtatgagtcatacaattattaccaccgagcgctacttattatttgaccggtcaattt 507
Db 36646 CTATATTTCTACGAGATATTACCATTTACAGCCGATATTGCATTCATTCGTGCTTTA 36587

Qy 508 ttaaatatatccaggcgagtgaggaacgctttatcaattatgtcagtttctctctt 567
Db 36586 ATAGATTTTGTGTTCTCTATGGGATTCATTTATTAAACATGATGTTATCTTTTCAGTT 36527

Qy 568 cctgcataatgtatgtattagcgcgctggtcttaggtcttttaaatcgtgcggca 627
Db 36526 CCCATGATTTATCGGTATATTCTTATGTGATATGGGTGTTGCTTAAACAAACAGCA 36467

Qy 628 caacaattgaatgtgtttttcttcaatgcgctcaaaagtatatattggttctactgacg 687
Db 36466 CCTCAGCTAAACGATTTACATTTATCTACATTCACGAGTAAAGAGTTTGATAGCAATCTTTATA 36407

Qy 688 ctctgatctcattcccttatctcttcat 717
Db 36406 TTGCTATTAGTTATTATGTTTTCCTGAT 36377

RESULT 13
AP002566/c
LOCUS      280900 bp      DNA      linear      BCT 07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 17/20.
ACCESSION  AP002566 BA000007
VERSION    AP002566.1 GI:13363930
KEYWORDS
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
ORGANISM    Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (sites)
AUTHORS     Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
            Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak

JOURNAL MEDLINE REFERENCE AUTHORS
20198780
2 (sites)
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MGL1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
3 (sites)
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
4 (sites)
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tober,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
5 (bases 1 to 280900)
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047)
genome project.
Location/Qualifiers
1..280900
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
complement(143..616)
/gene="ECs4456"
complement(143..616)
/gene="ECs4456"
/notes="similar to YIAI_ECOLI YIAI_ECOLI g1|2367245 percent
identity 98 in 157 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB37879.1"
/db_xref="GI:13363931"
/translation="MNRFTIADATKICGTCVACAVSHQENQDCAALSPDEFISRI
RVIKDHSWTTAVACHQCEPCANVCPVDAISRHHGHFVEQTRCIGCKSCMLACPPG
AMEVSSRRKKARAIKCDLCHWRHETGPACVEACTKALQCMDEKVKQRHRLRQOPV"
complement(860..1675)
/gene="ECs4457"
complement(860..1675)
/gene="ECs4457"
/notes="probable transcriptional regulator, weakly similar
to hypothetical transcriptional regulator YisR [Bacillus
subtilis] g1|3123306|sp|P40331 percent identity 24 in 276
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
```

```
/product-"putative transcriptional regulator"
/protein_id-"BAB37880.1"
/db_xref-"GI:13363932"
/translation-"MLEUSMTLPKIKVONGGLFISRGVGRHSARRLESWEIIFVEKRL
TIOEERIFLVDAGESLLPNRQHVGVDFPADLKFYLHFVKAPDNDPRWLTHIS
VPQTKVDPQALVALFRQFLSEQKHQRTPALEFIVLLILOQLTRDARLDEQAGAG
TTLAWKAQLIRTOYHRPLSTALLARELYCNDADYLGVRVRRVHLTUTEAHRQVRE
AEKLLISEARSLTEVAALCGFNDVGYFRQIFRKHHTGLTPPAWKKRRYSKEHINS"
1901. .3301
/gene-"ECs4458"
/length-1901
/gene-"ECs4458"
/notes-"probable permease, similar to hypothetical proteins
[Salmonella typhimurium] gi|7442781|pir|c65167 percent
identity 37 in 444 aa, also similar to transport proteins
(putative symporters) e.g. YicJ [Escherichia coli (K-12)]
gi|2851421|sp|p31435|YICJ_EC"
/codon_start-1
/transl_table-11
/evidence-not_experimental
/product-"putative permease"
/protein_id-"BAB37881.1"
/db_xref-"GI:13363933"
/translation-"MTSTPITTDIAKIKIDALSREKIGVGLDAGGVITCLIMN
FLTFYTVDFGLPALVGLFIALRFDIAISDPVMGVIAIDRTQSRWGRFRPWQWIAI
PIGIIGLFTVPDASMGVKIAWAFGVILLISVGYTAINVPYCARINTMTTRHEVIS
COSRFLVGVAGFLVSVGLPVMVALFOGNAARGQLGVGVLCIAVVMFLCCFFWV
RERVPLSTKGTFLREHLAQLRNNQDILLMLVMSFLINLVNFGGMYFTYVLOG
STGYTSLFTWMTVASIIGSVIVSPLTRFDYTKIYVYTNLLAALAVLWFLPSGA
YQTLWAVILGVNGVILGFTPLHFSLMFAFDYGYEWTNVRSSGMNFAFNLFKILAW
ASSAGISLFLIFVAYPGVQENOTASSJGIAAETLPALEFHLLLAGAIRCKLNNP
MMSRIATDLQRHVQp"
3312. .5282
/gene-"ECs4459"
/length-3312
/gene-"ECs4459"
/notes-"unknown, similar to hypothetical protein [Thermotoga
maritima] gi|7452109|pir|f72395 percent identity 37 in
635 aa"
/codon_start-1
/transl_table-11
/evidence-not_experimental
/product-"hypothetical protein"
/protein_id-"BAB37882.1"
/db_xref-"GI:13363934"
/translation-"WNISEVDLHKLTVSDPFLGQYQOQLVRDVVPIYQWDALNDRIPEA
EPHSAENFRIAAGLQEGEFYGMVQDSVAKWLEAVANSICQKPDALAELEKTADEVIE
LIASACEDGYLNTFTYVKAPEERWSNLAECHELICAGHLIEAGVAFQNTAKRRLIG
VVCRLADHSDVFGPDESKLHGYGHPETELALMRLYEVEPRYLALTYNVEQORGA
OPHYDOEYKRGQTSNHTYTGPAWVQDKAYSQHLPLAQOQTAIGHAVRFVYMTG
VAHLARLSDHDSKRDQCLRLWNNAROLYITGGISQSSGGEAFSSDYDLPNDTVYAE
SCASIGLMPFARMLEMEGDSQADVMERALYNTVIGGMALDGKHFYVNPLEVHPKS
LKENHIDYHVKPIRQRWFCACCPNIAARVLTSGHYLYTPREDALYINIIYAGNMEV
PVENGTLRVSGNTPWQOVTVAVESQPVRRHTLALRPMDCQTPQOILNGEVEQD
IRKGYLHITREMEGDTLNLTPMPVRRVYGNPLRVHVGAKVAIQRGPLVYCLEKADN
GESLHNLWLPDAPFTTPECKGLFSKILIQAPCYRYEDSNPQOQLPHHYDSAPAKRQ
TOTLTFIPWFSWANRGEEMRIWNEEKHCHP"
complement(5412. .6152)
/gene-"ECs4460"
/length-5412
/gene-"ECs4460"
/notes-"similar to YIAT_ECOLI gi|1790010 percent identity
99 in 246 aa (Conserved in E.coli K-12)"
/codon_start-1
/transl_table-11
/evidence-not_experimental
/product-"putative outer membrane protein"
/protein_id-"BAB37883.1"
/db_xref-"GI:13363935"
/translation-"MLINRNIVALPALPMASATASELSIGAGAAYNESPYRGYNENT
KAIPILISYEGDSFYRQTTIGFILSQSEKNELSLTASWMPLEFDPADNDNDYAMQQLDK
RDSAMAGVAYWHERRWGTWKASAAADVLDSNGWGLSVFKMOIGRLSLTPALGV
LYIDENFSDYIYGISESESRSSGLASYSAQDANWVPVYSLTAKYPIGEHVVLMASAGYS
```

```
ELPEITDSPMIDRNESFTFTVTGVSWRF"
6276. .7250
/gene-"ECs4461"
/length-6276
/gene-"ECs4461"
/notes-"similar to YIAU_ECOLI gi|1790011 percent identity
99 in 324 aa (Conserved in E.coli K-12)"
/codon_start-1
/transl_table-11
/evidence-not_experimental
/product-"putative transcriptional regulator lysR-type"
/protein_id-"BAB37884.1"
/db_xref-"GI:13363936"
/translation-"MTKLQKLYREKLIISVIAASENISHAATVGLIQANVSKYLADF
ESKVLGVDFDRTTQRLMLTFLPTALLPYINDMLDRNEQLNFIADYKHEKRGRTIYA
PTGIITLSKHVIDIKIDIGDITLSLKTCLNLENAFYEGVEFPDDCDVLISYAPPKDE
SLVASFITQYAVTAYASQRYLEKHPISRPELEHHSCILIDSMIDANINWRPNVAGS
KEVRDYRVKGYVCDNTQSALARNHLGIVFAPDKSVQSDLDQDGLVPCFQHPYEWV
LDLVAIFRKREYQPMRVQYVLDMLREIRHQLAQSQLRPEQAASED"
complement(7247. .8383)
/gene-"ECs4462"
/length-7247
/gene-"ECs4462"
/notes-"similar to YIAV_ECOLI gi|1790012 percent identity
99 in 378 aa, similar to YIBH_ECOLI gi|1790024 percent
identity 76 in 378 aa (Conserved in E.coli K-12)"
/codon_start-1
/transl_table-11
/evidence-not_experimental
/product-"putative membrane protein"
/protein_id-"BAB37885.1"
/db_xref-"GI:13363937"
/translation-"MDLLIILTYVAFWAFWPKIPVKNKWTIPTAALGGIFIVSGLI
LLMNYNHPYTFKAQKAVISIPVQVTVIEVTDKNTLTKKEVLEFRIDPTRYQAR
VDRLMADIVTAEHKORALGAELDEMAANTQKAKATRDKPAKQYQRYARGSOAKVNPFS
ERDIDVARQNYLAQESVSKSAABEQIQSOLDSLVLGESHQSLASLKAQLAEAKYNLE"
Query Match 15.1%; Score 118; DB 1; Length 280900;
Best Local Similarity 52.0%; Pred. No. 1.7e-18;
Matches 265; Conservative 0; Mismatches 245; Indels 0; Gaps 0;
QY 208 tggtaggttagtcaccggagaggtgattatggttttttaattgggttttgcggcg 267
Db 129813 TTGACAGGAATAGCGCTAAAGAAATAAGTATTGGTTTTCATTGGGTATTACATTTACT 129754
QY 268 gtcccttttggccgttgatggcggttcttgccttgcattcattcagtgggcgaca 327
Db 129753 ATTCTTTTGGGCAATAGATGCGGCTGGACAGATTATTGATATCTTAAGAGGCTCAACA 129694
QY 328 atgggtgacgattcattcattcattcattcattcattcattcattcattcattcattc 387
Db 129693 ATATCTTCAATTTTAAACCGCTCCATAAGTATTCATCTTCTATCCTGCGGTATCTTG 129634
QY 388 agccagttttgtgtgtattttcttataagcggcgccatggagtttataataacatt 447
Db 129633 TACCAATTTATCTGTGATCTTTGTTATTTCATGTTGGGATACAAGCATCCTAGATAAG 129574
QY 448 ctgttagtgatcatcatatattaccaccagggtcgctacttattatttggaccggcaatt 507
Db 129573 CTATATTTATCTACGAGATATTACCATTTACAAGCGGATATTGCATTCAATCGTCTTTA 129514
QY 508 ttaaaatataatccaggcagagtggaacgctttatcaattatggtcagttcttctt 567
Db 129513 ATAGATTTTCTTCTATGGGATTCATTTATTAACATGATCTTATCATTTTCAGTT 129454
QY 568 cctgccataaatgtagtgattagccgagatggtttaggtcttttaaatcggcgcga 627
Db 129453 CCCATGATTATCGGTATATTCTTATGTGATATGGGTTTGGGTTTCTTAACAAACAGCA 129394
QY 628 caacaattgaatgtgttttcttctcaatgcgcgctcaaaagtatatattggttctactgacg 687
Db 129393 CCTCAGCTAAACGTATTCACATTATCCTACTGCCAGTANAAAGTTTGTAGCAATCTTTATA 129334
```

```

QY 688 ctctgatctcattcccttatctcttcat 717
      ||| | | | | | | | | | | |
Db 129333 TTGCTATTAGTTATTCATGTTTTCCTGAT 129304

RESULT 14
AF200363
LOCUS DEFINITION
AF200363 37889 bp DNA linear BCT 27-MAR-2001
Escherichia coli ler (ler), EscR (escR), EscS (escS), EscT (escT),
EscU (escU), EscD (escD), EscC (escC), EscQ (escQ), SepZ (sepZ),
EscV (escV), EscN (escN), SepQ (sepQ), Tir (tir), Cest (cest),
IntimB (eae), EscD (escD), SepL (sepL), EspA (espA), EspD (espD),
EspB (espB), EscF (escF), EspF (espF), and putative transposase
genes, complete cds; Lifa (lifa) gene, partial cds; and unknown
genes.
ACCESSION AF200363
VERSION AF200363.1 GI:13447694
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 37889)
Agin,T.S., Cantey,J.R., Boedeker,E.C. and Wolf,M.K.
Characterization of the eaeA gene from rabbit enteropathogenic
Escherichia coli strain RDEC-1 and comparison to other eaeA genes
from bacteria that cause attaching-effacing lesions
FEMS Microbiol. Lett. 144 (2-3), 249-258 (1996)
97055784
8900070
2 (bases 1 to 37889)
Elliott,S.J., Wainwright,L.A., McDaniel,T.K., Jarvis,K.G.,
Deng,Y.K., Lal,L.C., McNamara,B.P., Donnenberg,M.S. and Kaper,J.B.
The complete sequence of the locus of enterocyte effacement (LEE)
from enteropathogenic Escherichia coli E2348/69
Mol. Microbiol. 28 (1), 1-4 (1998)
98254123
3 (bases 1 to 37889)
Zhu,C., Agin,T.S., Elliott,S.J., Johnson,L.A., Thate,T.E.,
Kaper,J.B. and Boedeker,E.C.
Complete Nucleotide Sequence and Analysis of the Locus of
Enterocyte Effacement from Rabbit Diarrheagenic Escherichia coli
RDEC-1
Infect. Immun. 69 (4), 2107-2115 (2001)
21153569
PUBMED 11254564
4 (bases 1 to 37889)
Boedeker,E.C., Zhu,C., Elliott,S.J., Tonia,T.S., Johnson,L.A.,
Thate,T.E. and Kaper,J.B.
Direct Submission
Submitted (01-NOV-1999) University of Maryland School of Medicine,
Center for Vaccine Development, 685 West Baltimore St., Baltimore,
MD 21201, USA
FEATURES
Location/Qualifiers
1..37889
/organism="Escherichia coli"
/strain="RDEC-1"
/serotype="O15:H-"
/db_xref="taxon:562"
/clone="pDK5"
/note="Rabbit EPEC, REPEC"
complement(<1..396)
/organism="Escherichia coli"
/strain="RDEC-1"
/db_xref="taxon:562"
/clone="pDK5"
1032..35232
/note="RDEC-1 locus of enterocyte effacement (LEE)
pathogenicity island core region; similar to sequences
presented in GenBank Accession Numbers AF022236 and
AF071034"
complement(1209..2027)

source
misc_feature
CDS
gene
6698..7351
/gene="escr"
6698..7351

```

```

/note="rOrf1"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK26694.1"
/db_xref="GI:13447695"
/translation="MKRILFAFFIAMFVTPIALILFVFDYKKSPQVSLSFQKL
PNTVLSLTYNKREDDFLYATONSPAOLKKTARSGLIIGRAPLPIISDAETIEHQGN
IFAADVDEKTSLEFFFTYTKDMHISLRNKIQLEKFNKKNRPFEGLAWKAGRMILFVAKE
RRPTGMFAQLSPDLLSAKQVTIPEELNDIHVKDISGLDNFNESMILSDSKLLKF
NLTEMSFVEMMDTKGNHSLTSDLLQPEGIVTUSDESIYVASEPDILAKFVPNK"
complement(2172..3368)
/note="rOrf2"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK26695.1"
/db_xref="GI:13447696"
/translation="MISVVEFIIINETSIMINGLNNSASLVDAAIRINSDFKKQ
WDMSCAEKLLKVLSPGLNPTYRSEQTFFQELTVLEPVSAPNELGRIYANFADG
SSLRISVTNSELVEAEIRTPDNEKILMLLESNEQNRLLQSLPINLHMPYIOVHRALSK
MDLTDHKSMMHLLSFTSKLSATLI PHNTOTDPLSGPTPSSMMDTFRGLGNAKLSLN
GVDIPVDAQKLLDALGLKTHSSLARVINNGISRHHAKQIARESSGSDKOKAEVVE
FLCHPEAATATCSAFYQSFNVYPALMLTHTRISQAREYNVRSIDVPNACINISISQSP
DGSIHVASHTGILIMAPEDRPNELGMLTNRTSYEVPOGVKCEIDEMVRTLQPRYGASE
TYLKNK"
4508..4897
/gene="ler"
4508..4897
/gene="ler"
/codon_start=1
/transl_table=11
/product="ler"
/protein_id="AAK26696.1"
/db_xref="GI:13447697"
/translation="MRRLEIMNMTNHTTSPVLIQLEIEVLQEQEAKRLREQEIQSV
IESIQKITYNITLQELGYTNVPDDGLARRNSKGYRNEEGQWTSVGVRQPRWLK
EALLNGMKREDFLVKDEDEIIPLKNI"
4912..5130
/note="Orf2"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK26697.1"
/db_xref="GI:13447698"
/translation="MITTELEDEIIKNKEANIFIEKINDKKNETHKMKHPLDKVT
YNEAKELIACDAAIRIIEIMLRINK"
5454..6053
/note="Orf4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK26698.1"
/db_xref="GI:13447699"
/translation="MTILNKIDYNYKLYNPIMVHDEWLGDLTGVMQVSRHLRES
SSTENQLNKILRQEIQKISGVLSINKEGFLYQSIGIRLLALSALFEIQCPDYI
FSRLIRETLIREIGYQNVKQSFYWGOGKPEYGEERFCSELIKYGAGNLEWLFSDN
PLMTIVKYLPRKSGEIKRPTHINDLFENRLNKLILPYETL"
5998..6693
/note="Orf5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK26699.1"
/db_xref="GI:13447700"
/translation="MIVFLTDLIKFYCLMKLYEYKNIKIDLVLTEIIPEDKLEIIQ
SDDIVKLARKKTCHELLRARKSKKLTESRKTARKIMAMREIRKNNKIKLEKEVN
KSIKWKDIQDIELVLMQDMNKVHSLTNALHSDTSRRINDDLLNEVRETLNIN
NIVGAIKITKNPDIKLDPGGEANNIQLINDADAPLNKIITENEYMRITLDPLEQINILL
NSFKENYLSIIQE"
6698..7351
/gene="escr"
6698..7351

```

```
/gene="escr"
/codon_start=1
/transl_table=11
/product="Escr"
/protein_id="AAK26700.1"
/db_xref="GI:13447701"
/translation="MSQLMTIGSQPIFLIIVFLLSLLPIFVVIGTSFLKISIVLGL
KNAIGIQVPPNMAITSVSLITMPSIILQINDNISQEPINYTDSDFOKVDEKI
LSPYGFLENTKENVFERAAOKKLGNETILKDSLFILPAPFMGOLAAEFKIG
FLYIPFPIADLIISNILLALGMMWSPYISIPFKILLFILVGGWOKLFEFLVYN"
7351..7620
/gene="escs"
7351..7620
/gene="escs"
/codon_start=1
/transl_table=11
/product="Escs"
/protein_id="AAK26701.1"
/db_xref="GI:13447702"
/translation="METGFEVQLCVQTLWIIFILSLPTVIAASVVGIIISLVQAITQL
QDQTLPLFLKIIIAVFATLALTYHMGTTIINFSSIIIFEMIPRVNG"
7620..8396
/gene="escT"
7620..8396
/gene="escT"
/codon_start=1
/transl_table=11
/product="EscT"
/protein_id="AAK26702.1"
/db_xref="GI:13447703"
/translation="MNEIMTVIVSSFFCILRPLGMFIILPSTGVLLSNFIRNSIMI
AFTLPITVENYTFSEKLPSCI FOLTGIVLKEISIGFFIGLSFTILFWAIDAAGQIIDT
LRGSISSINPNSISDSSITGVILYQFIVFVHGGIQSVLDKLYLSEILPLOAD
IAFNALRDLFSLWDSFIKLSFSIPMIGIIFLCDMGFGFLNKTAPOLVNFTLSLP
VKSLTAIFILLVIHVFPDITANIHSIDIIRSLPSMINE"
8389..9426
/gene="escU"
8389..9426
/gene="escU"
/codon_start=1
/transl_table=11
/product="EscU"
/protein_id="AAK26703.1"
/db_xref="GI:13447704"
/translation="MSEKTEKTPPKRLDKRKGDTKSEEVMAAVOSLILFFSFLY
GASFFVEVVELVNTTIDSLNRPFLYAIREILGAVLNIFLLYILPISLIVFGVTGTG
SQTGFIFAVBKIPSAQKISVKNLKNFVSFKSIFELKSVFKLVIITLIFFMGHSY
ANEAFNTRLNAYOTLVVAVFFVFLWGVLFVGFVDFWFKRHEGLKMKMSKD
EVKRAKDTQGNPEIKGERRLHSEIQSGSLANNIKKSVIVKNPHTHAICLYKIGE
TPLVIETGDKAKALQIILKLAELYDIPYIEDIPLARSLYKNIHKQYITDFEPVA
QLIRTAIDLDT"
complement(9423..9881)
/notes="orf3"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK26704.1"
/db_xref="GI:13447705"
/translation="MKIILISILISNSCYASDCFEITGKAYNTDPLILKAJWNSK
NKGINKINKNTYDIGINQINSHLDLLSKFNISEDLLNDACINISVAGYILASN
IKSRGNTWDVAGYAGYFNTPNAVELRQYAKKIYKTYTKLNNEQIID"

Query Match      14.6%; Score 113.8; DB 1; Length 37889;
Best Local Similarity 52.7%; Pred. No. 2.1e-17;
Matches 247; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

QY 208 ttggttaggttagtcaccggaggtgattggttttttaattgggttttgcgcgcg 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7821 TTCACAGCATAGTGCTAAAGAAATAAGTATAGGCTTTTCATTGGGTATCATTTACT 7880
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 268 gttcccttttggcggtgattgagcggtgttctgcttgatacttaccgtggcgcgaca 327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7881 ATTCTTTTGGGCAATAGATCGCGCTGGCGAGATATTGATACATTAAAGAGGTTCAACA 7940
```

```
QY 328 atggggtacgatattcaattctacaatagaaagctgaaccctcaactttttggctgtcttc 387
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7941 ATATCTTCAATTTTAAACCCGCTATAAGTGATTTCATCTTCTATCACTGGTGTATTTTTA 8000
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 388 agccaggttttggtgtgtattttttttataaagcgcggtgagtttataataaacattc 447
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8001 TACCAATTTATCACTGTCATCTTTGTTATTATCATGTGGGATACAAACGCTTCTGCATAAG 8060
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 448 ctgatatgagtcatacaatatttaccaccagggtgclacttattattttgaccggcaattc 507
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8061 TTGTATTATCTACGAGATATTACCATTAACAAGCGGATATTGCATTCAATCGTCTTTTA 8120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 508 ttaaaatataccaggcagagtgaggagacgttttatacaattatgctgagtttctctctt 567
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8121 GTAGATTTTATTTTCGCTATGGGATTTCATTATATANAACGTATTCATTTTCATTT 8180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 568 cctgccataatatgtatgtattagcgcgatcgtgtttaggtcttctttaaactcggtcgcca 627
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8181 CCCATGATTATCGGTATCTTCTTATGTGATATGGGTTTGGGTTTCTCAATAAAACAGCG 8240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 628 caacaatgaatgtgttttcttctcctcaatgcgcgtcaaaaagtatatgg 676
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8241 CCTCAGCTAAACGTATTTCACATTATCACTGCCGCTAAAAAAGTTTGATAG 8289
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 15

ECO277443

LOCUS

DEFINITION

ECO277443 60073 bp DNA linear BCT 05-DEC-2001

Escherichia coli DNA for locus of enterocyte effacement II (LEE

II).

AJ277443

AJ277443.1 GI:17384625

cadC gene; CadC protein; cad gene; CaeD protein; cest gene; Cest

protein; eae gene; eaeA gene; eaeB gene; eaf-11fA-tox gene;

Efa1-LifA-Tox protein; ent gene; Ent protein; escC gene; EscC

protein; escD gene; EscD protein; escF gene; EscF protein; escJ

gene; EscJ protein; escN gene; EscN protein; escR gene; EscR

protein; eacS gene; EacS protein; eacT gene; EacT protein; escU

gene; EscU protein; escV gene; EscV protein; espA gene; EspA

protein; espB gene; EspB protein; espD gene; EspD protein; espE

gene; EspE protein; espF gene; EspF protein; int gene; Integrase;

intimin; ler gene; Ler protein; orfU gene; p4 gene; sepA gene; sepB

gene; sepC gene; sepD gene; sepF gene; sepG gene; sepQ gene; scpH

gene; sepI gene; sepL gene; SepL protein; sepQ gene; SepQ protein;

sepZ gene; SepZ protein; st01 gene; ST01 protein; st02 gene; ST02

protein; st03 gene; ST03 protein; st04 gene; ST04 protein; st06

gene; ST06 protein; st07 gene; ST07 protein; st08 gene; ST08

protein; st13 gene; ST13 protein; st14 gene; ST14 protein; st15

gene; ST15 protein; st18 gene; ST18 protein; st20 gene; ST20

protein; st22 gene; ST22 protein; st25 gene; ST25 protein; st26

gene; ST26 protein; st28 gene; ST28 protein; st29 gene; ST29

protein; st30 gene; ST30 protein; st39 gene; ST39 protein; st41

gene; ST41 protein; st43 gene; ST43 protein; st44 gene; ST44

protein; st46 gene; ST46 protein; st47 gene; ST47 protein; st48

gene; ST48 protein; st50 gene; ST50 protein; st51 gene; ST51

protein; st52 gene; ST52 protein; st53 gene; ST53 protein; st54

gene; ST54 protein; st55 gene; ST55 protein; st56 gene; ST56

protein; st57 gene; ST57 protein; tir gene; transposase.

Escherichia coli.

Escherichia coli

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Escherichia.

1 (sites)

Benkel, P. and Chakraborty, T.

Genetic organisation and sequence of the LEE II locus in Shiga

toxin-producing Escherichia coli

Unpublished

2 (bases 1 to 60073)

Benkel, P.

Direct Submission

Submitted (19-APR-2000) Benkel P., Justus-Liebig-Universitaet

Giessen, Institut fuer Medizinische Mikrobiologie, Frankfurt


```

/translate-11
/product-"EsCR protein"
/protein_id-"CAC81847.1"
/db_xref-"GI:17384635"
/translation-"MSQMTIGSQPIFLIIVFLLSLPFIWVIGTSFLKISIVLGIL
KNALGIQOVPPNMLTSLILTMFIMSLIILQINDNISOEPINYTDSDFOKVDEKI
LSPYRGPLENTERNEVEFFERAQKKLGNETILKKDSLFIILLPFTMGQLEAAFKIG
FLLYLPFIAIDLIISNIIILALGMHVPVTISIPFKILLFILVGGWOKLFEFLLVN"
8455..8724

gene
Query Match      14.6%; Score 113.8; DB 1; Length 60073;
Best Local Similarity 52.7%; Pred. No. 2e-17;
Matches 247; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

QY 208 tggtaggttagtcaccggagagtgattggttttttaattgggttttggcgcg 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8925 TTGACAGGCATAGTGTAAAGAAATAAGTATAGGCTTTTCATGGGTATCATTTACT 8984

QY 268 gtcccttttggcgcttgatgctggcggtttctctgtgatactttacgtggcgcgaca 327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8985 ATTCTTTTGGGCAATAGATGCGGCTGGCGAGATTATTGATACCTTTAAGAGGTTCAACA 9044

QY 328 atggggtacgatatccattctacaatagaagctgaacctcaacttttggcttgccttttc 387
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9045 ATATCTTCAATTTTAAACCCGCTATAAGTGATTTCATCTTCTATCACCTGCTGTTATTTA 9104

QY 388 agccagtttttgtgtgtattttcttataagcgcggaatggagtttatataaaacatt 447
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9105 TACCAATTATCACCTGTGATCTTTGTTATTTCATGCTGGGATACAAAGCGTTCGGATAAG 9164

QY 448 ctgtatgagtcataccaattattaccacggcggtactttattatttaccggcaattt 507
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9165 TTGTATTATCTTACGAGATATTACCATTTACAGCGGATATTGCATTCATCTGCTTTA 9224

QY 508 tlaaaatataccaggcagagtggaagacgctttatcaattatgtgcagtttctctctt 567
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9225 GTAGATTTTATTATTTCCCTATGGGATTCATTTATTAACTGATGTTATCATTTTCAATT 9284

QY 568 cctgccataatagtatgggtatagccgatctggttttaggtctttttaaatacggtcgga 627
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9285 CCCATGATTATCGGTATCTTCTTATGTGATATGGGTTTGGGTTTCTCAATAAACACGCG 9344

QY 628 caacaattgaatgtgtttttcttcaatgccgcgtcaaaagtatatgg 676
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9345 CCTCAGCTAAAGCTATTACATTTATCAGTCTGCGGTAAAGTTTGATAG 9393
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 00:47:00 ; Search time 236.42 Seconds
(without alignments)
5657.205 Million cell updates/sec

Title: US-09-809-524B-1
Perfect score: 779
Sequence: 1 atggcacaacaggtaaatga.....gactgggttcacatctgtatg 779

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	779	100.0	779	22	Salmonella dublin
2	763	97.9	779	22	Salmonella typhimur
3	750.6	96.4	13417	17	Partial sequence o
C 4	207.4	26.6	297	17	Virulence factor w
5	60.2	7.7	801	21	Bordetella pertuss
6	60.2	7.7	35026	21	Bordetella pertuss
C 7	52.2	6.7	184	17	Virulence factor w
8	42.4	5.4	6866	24	Human immune syste
9	42.4	5.4	16914	24	Human gene regulat

10	41.8	5.4	6102	24	AAS61356	Human gene regulat
11	41.2	5.3	16842	22	AAS46411	Tumour suppressor
12	41.2	5.3	16842	24	AAS61335	Human gene regulat
13	40.6	5.2	17538	24	ABL33157	Human immune syste
14	40	5.1	26997	22	AAS46748	Tumour suppressor
15	39.8	5.1	5311	24	ABL33019	Human immune syste
16	39.6	5.1	707	22	AAL24056	Human breast cance
17	39.6	5.1	73334	24	ABL34125	Human immune syste
18	39.4	5.1	9741	24	ABL33323	Human immune syste
C 19	39	5.0	5059	20	AAH84332	Stealth virus nucl
20	39	5.0	17738	24	ABL33539	Human immune syste
21	39	5.0	19576	24	AAS61258	Human gene regulat
22	39	5.0	34769	22	AAS46775	Tumour suppressor
23	38.8	5.0	8136	24	ABL32555	Human immune syste
24	38.8	5.0	8143	24	ABL33897	Human immune syste
25	38.8	5.0	13814	24	ABL33193	Human immune syste
26	38.6	5.0	5296	24	ABL33285	Human immune syste
27	38.6	5.0	53585	20	AAH20251	Borrelia burgdorfe
28	38.4	4.9	6782	24	ABL32776	Human immune syste
29	38.4	4.9	8349	24	ABL33789	Human immune syste
30	38.2	4.9	6047	24	ABL33459	Human immune syste
31	38.2	4.9	6118	24	ABL33031	Human immune syste
32	38.2	4.9	9524	24	ABL32839	Human immune syste
33	38	4.9	6558	22	AAS46549	Tumour suppressor
34	38	4.9	15224	24	AAS61262	Human gene regulat
35	37.8	4.9	6189	22	AAS46599	Tumour suppressor
36	37.8	4.9	6189	24	ABL34592	Human metastasis a
37	37.8	4.9	13511	24	ABL32281	Human immune syste
38	37.6	4.8	5574	24	AAS63338	Chemically pretrea
39	37.6	4.8	6381	24	ABL32967	Human immune syste
40	37.6	4.8	6381	24	ABL34519	Human metastasis a
41	37.6	4.8	7134	24	ABL32482	Human immune syste
42	37.4	4.8	11047	22	AAS45480	Chemically pretrea
43	37.4	4.8	11047	24	ABL33985	Human immune syste
44	37.4	4.8	15649	22	AAS45396	Chemically pretrea
45	37.4	4.8	20510	23	ABL19710	Drosophila melanog

ALIGNMENTS

RESULT 1

AAD20626
ID AAD20626 standard; DNA: 779 BP.

XX AC AAD20626;

XX DT 03-JAN-2002 (first entry)

XX DE Salmonella dublin ssat gene.

XX ssat; antibacterial; vaccine; non-rodent animal; immunity; mortality;

XX diarrhoea; milk production; bacterial infection; ds.

XX OS Salmonella dublin.

XX PN WO200170247-A2.

XX PD 27-SEP-2001.

XX PF 13-MAR-2001; 2001WO-US08042.

XX PR 17-MAR-2000; 2000US-190178P.

XX PA (PHAA) PHARMACIA & UPJOHN.

XX PI Lowery DE, Kennedy MJ;

XX DR WPI; 2001-639093/73.

XX PT Vaccine composition useful for conferring protective immunity in a

PT non-rodent animal, comprises first attenuated, non-reverting mutant

PT Salmonella bacterium having two or more inactivated genes within SPI2


```
OY 121 aatggcgcttattgacactacattccattccatattaccatcattaccagcagaagatt 180
|||||
DB 121 aatggcgcttattgacactacattccattccatattaccatcattaccagcagaagatt 180
|||||
OY 181 atgatacatattgtaaaagattacagtttaggttaggttagtcacccggagaggtgattatt 240
|||||
DB 181 atgatacatattgtaaaagattacagtttaggttaggttagtcacccggagaggtgattatt 240
|||||
OY 241 ggttttttaattgggtttttagtcggcggttcccttttggccggttgatcgcggtgttt 300
|||||
DB 241 ggtttttcaattgggtttttagtcggcggttcccttttggccggttgatcgcggtgttt 300
|||||
OY 301 ctgcttgatacttttaagtcggcgacaaatgggtacgataatcattacataagaagct 360
|||||
DB 301 ctgcttgatacttttaagtcggcgacaaatgggtacgataatcattacataagaagct 360
|||||
OY 361 gaaacatcacttttggcttcttccagccagtttttgggtgtgttattttcttataagc 420
|||||
DB 361 gaaacatcacttttggcttcttccagccagtttttgggtgtgttattttcttataagc 420
|||||
OY 421 ggcggcatggagtttatataaacaattctgtatgagtcataataattaccaccaggg 480
|||||
DB 421 ggcggcatggagtttatataaacaattctgtatgagtcataataattaccaccaggg 480
|||||
OY 481 cgtactttatttggacggcgaatttttaaaatatatccagcgagtggaagacgctt 540
|||||
DB 481 cgtactttatttggacggcgaatttttaaaatatatccagcgagtggaagacgctt 540
|||||
OY 541 tatcaattatgtgacagttctctctctctccgcataataatgtatggttagcgcgactg 600
|||||
DB 541 tatcaattatgtgacagttctctctctctccgcataataatgtatggttagcgcgactg 600
|||||
OY 501 gotttaggtcttttaaatcggctcgccacaaatgaatgtgttttcttccatgcgcg 660
|||||
DB 501 gotttaggtcttttaaatcggctcgccacaaatgaatgtgttttcttccatgcgcg 660
|||||
OY 561 ctcaaaagtataattggttctactgacgctctgactctcattcccttattcgtcttccatcac 720
|||||
DB 561 ctcaaaagtataattggttctactgacgctctgactctcattcccttattcgtcttccatcac 720
|||||
OY 721 tatttggtgaaagcgataaatttatatttatctataaagactggtttccatctgtatg 779
|||||
DB 721 tatttggtgaaagcgataaatttatatttatctataaagactggtttccatctgtatg 779
|||||

RESULT 3
AAT09224
ID AAT09224 standard; DNA; 13417 BP.
XX
AC AAT09224;
XX
DT 07-JAN-1997 (first entry)
XX
DE Partial sequence of virulence gene cluster 2 of S. typhimurium.
XX
KW Mutant; adaptation; virulence factor; identification; screening;
KW vaccine; drugs; infection; treatment; ds.
XX
OS Salmonella typhimurium.
XX
FH Key Location/Qualifiers
FT CDS 262
FT /tag= a
FT /note= "Putative start site of yscJ gene."
FT CDS 237
FT /tag= b
FT /note= "Putative start site of yscJ gene."
FT misc_feature 936..937
FT /tag= c
FT /label= Transposon insertion site (P11H11).
FT misc_feature 1074..1075
FT /tag= d
```

```
FT CDS
1251 /label= Transposon insertion site (P11D10).
/*tag= e
/*note= "Putative start site of lcrE gene."
3646
/*tag= f
/*note= "Putative start site of lcrD gene."
4121..4122
/*tag= g
/*note= "Transposon insertion site (P2D6)."
4579..4580
/*tag= h
/*label= Transposon insertion site (P11C3).
5622
/*tag= i
/*note= "Putative end site of lcrD gene."
5627
/*tag= j
/*note= "Putative start site of yscN gene."
5794
/*tag= k
/*note= "Putative start site of yscN gene."
6548
/*tag= l
/*note= "Putative start site of yscN gene."
6910
/*tag= m
/*note= "Putative end site of yscN gene."
7226
/*tag= n
/*note= "Putative end site of yscQ gene."
7271
/*tag= o
/*note= "Putative start site of yscP gene."
7626
/*tag= p
/*note= "Putative start site of yscQ gene."
7645
/*tag= q
/*note= "Putative end site of yscP gene."
7710
/*tag= r
/*note= "Putative start site of yscQ gene."
8595
/*tag= s
/*note= "Putative end site of yscQ gene."
8662
/*tag= t
/*note= "Putative start site of yscR gene."
9309
/*tag= u
/*note= "Putative end site of yscR gene."
9306
/*tag= v
/*note= "Putative start site of yscS gene."
9572
/*tag= w
/*note= "Putative end site of yscS gene."
9573
/*tag= x
/*note= "Putative start site of yscT gene."
9801..9802
/*tag= y
/*label= Transposon insertion site (P9B7).
10351
/*tag= z
/*note= "Putative end site of yscT gene."
10349
/*tag= i
/*note= "Putative start site of yscU gene."
11059..11060
/*tag= i
/*label= Transposon insertion site (P12F5).
```


XX PS Claim 32; Figure 6; 131pp; English.

XX CC A method for identifying a microorganism having a reduced adaptation

CC to a particular environment comprising the steps of: (1) providing a

CC plurality of microorganisms each of which is independently mutated by

CC the insertional inactivation of a gene with a nucleic acid comprising

CC a unique marker sequence so that each mutant contains a different

CC marker sequence, or clones of the said microorganism; (2) providing

CC individually a stored sample of each mutant produced by step (1) and

CC providing individually stored nucleic acid comprising the unique

CC marker sequence from each individual mutant; (3) introducing a

CC plurality of mutants produced by step (1) into the said particular

CC environment and allowing those microorganisms which are able to do so

CC to grow in the said environment; (4) retrieving microorganisms from

CC the said environment or a selected part thereof and isolating the

CC nucleic acid from the retrieved microorganisms; (5) comparing any

CC marker sequences in the nucleic acid isolated in step (4) to the

CC unique marker sequence of each individual mutant stored as in step

CC (2); and (6) selecting an individual mutant which does not contain any

CC of the marker sequences as isolated in step (4). The products and

CC methods can be used for identifying virulence genes in microorganisms.

CC The mutant microorganisms can be used in vaccines or to screen for

CC drugs which reduce virulence or compounds useful for preventing,

CC ameliorating or treating infections in animals or plants. This

CC virulence factor sequence was designated p9B7_1.I.

XX SQ Sequence 297 BP; 99 A; 67 C; 55 G; 76 T; 0 other;

Query Match 26.6%; Score 207.4; DB 17; Length 297;

Best Local Similarity 97.2%; Pred. No. 6.5e-48;

Matches 211; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atggcacaacagtgaaatgagtggtattgctggtgctggtgcttttattcgaccatta 60

DB 224 ATGGCACACACAGTAATGAGTGGCTATTTCATTTGGCTGTGGCTTTTATTCGACCATG 165

QY 61 agcctttctttattactccctattataaagtgaggcagtttagggcgctctttacgt 120

DB 164 AGCCTTTCTTTATTACTTCCCTATTATAAAGTGGCAGTTTAGGGCGCGACTTTTACGT 105

QY 121 aatggcgctgtatgtcacttacccttccctattaccatcattaccagcagaagatt 180

DB 104 AATGGCGTGTATGTGCTACCTTCCGATATTACCAATCATTTACCAGCAGAGATT 45

QY 181 atgatgatattgggtaagattacagttggttagggt 217

DB 44 ATGATGCATATTGGTAAAGATTACAGTTGGGTAGGTT 8

RESULT 5

AAA64865

ID AAA64865 standard; DNA; 801 BP.

XX AC AAA64865;

XX 02-FEB-2001 (first entry)

XX Bordetella pertussis class I gene bscT coding sequence.

DE bscT; bacterial infection; anti-bacterial; vaccine; whooping cough;

KW type III secretion system; virulence factor; pathogenicity island; ss.

XX Bordetella pertussis.

XX Key Location/Qualifiers

FT 1..801

FT CDS /tag= "BscT"

FT /product= "BscT"

XX PN W0300037493-A2.

XX

PD 29-JUN-2000.

XX 21-DEC-1999; 99WO-EP10297.

XX 21-DEC-1998; 98GB-0028217.

XX (ULBR) UNIV LIBRE BRUXELLES.

PI Bollen A, Fauconnier A, Godfroid E;

XX WPI; 2000-452178/39.

DR P-PSDB; AAB14127.

XX Novel polypeptides derived from Bordetella pertussis, useful for

PT treating and diagnosing Bordetella infection -

XX Example 2; Pages 115-116; 165pp; English.

XX Bordetella pertussis possesses a type III secretion system. Type III

CC secretion systems allow bacteria to target virulence factors directly at

CC host cells. The present sequence is the bscT coding sequence of B.

CC pertussis. The present sequence is a Class I type gene and encodes a

CC protein involved in the type III secretion system of B. pertussis i.e.

CC a Bordetella pathogenicity protein. The present sequence is located

CC within a pathogenicity island (see AAB64890). A pathogenicity island is

CC a compact, distinct genetic unit carrying virulence genes. The protein

CC encoded by the present sequence may be used to treat or diagnose B.

CC pertussis infection, e.g. as a vaccine. Whooping cough is a disease

CC caused by infection by B. pertussis.

XX SQ Sequence 801 BP; 105 A; 244 C; 259 G; 193 T; 0 other;

Query Match 7.7%; Score 60.2; DB 21; Length 801;

Best Local Similarity 47.9%; Pred. No. 7e-07;

Matches 235; Conservative 0; Mismatches 253; Indels 3; Gaps 2;

QY 210 gttagggttagtcaccggagaggtgattattggttttttaattggttggcgggcggt 269

DB 219 gctggcgctgctggccaaaggagcgatggtggcgatgcttccctgggtggcgctgctt 278

QY 270 tcccttttggcgcttgatattggcggtttctgcttgatatacttttaccgtggcgcgcaat 329

DB 279 gccattctggatcttcggagccatcggtcttcgcatagacaacacggcgcgcgacct 338

QY 330 ggtgacgatattcaattctacaatagagctgaaacctcacttttttggctgcttttcag 389

DB 339 gggcgctatctcaaccccgccacggcgcaacgattcgctcgccatggcgatctcttcaa 398

QY 390 ccagtttttggctgtttatttttttataagcgcgcgcgatgagtttataataaacattct 449

DB 399 tctgggattcattggttcttctcagcgcgcggttgcggttgcggttgcgcatgct 458

QY 450 gtatga-gtcatatcaattattaccacagcggtgactttattatttaccggcgcaatttt 508

DB 459 gtatgacagcttcgggttggaacatctggcggtggtggcggtccatgcccgacacagg 518

QY 509 taaaatat--atccagcgagagtggaacacgctttatcaattatgtgtcagtttctctct 566

DB 519 cgcggtgcgagatgctggaccagttcagtggttgcgcgcggtgctgctgctgctc 578

QY 567 tctgcccataatagtattggtattaccacagcggtgactttattatttaccggcgcaatttt 626

DB 579 gccggccatcgtggtccatgttctcgtgagctggcggtggtggtggtggtggtggtggt 638

QY 627 acaacaattgaatgtgttttttcttctcaatgccgctcaaaaagtatatattggttctactgac 686

DB 639 gcctcaactgcaggtgttcttctcgtggtctgctgctgctgctgctgctgctgctgct 698

QY 687 gctcctgatct 697

DB 699 gctggtgctgt 709

```

RESULT 6
AAA64890
ID AAA64890 standard; DNA; 35026 BP.
XX
AC AAA64890;
XX
DT 02-FEB-2001 (first entry)
XX
DE Bordetella pertussis pathogenicity island coding sequence.
XX
KW Bacterial infection; anti-bacterial; vaccine; whooping cough;
KW type III secretion system; virulence factor; pathogenicity island; ss.
XX
OS Bordetella pertussis.
XX
FH Key Location/Qualifiers
FT CDS 711..2024
FT /*tag= a
FT /product= "Protein # 1 (AAB14147)"
FT complement (2055..3590)
FT /*tag= b
FT /product= "Protein # 2 (AAB14148)"
FT 4220..4696
FT /*tag= c
FT /product= "Protein # 3 (AAB14149)"
FT complement (4998..5948)
FT /*tag= d
FT /product= "Protein # 4 (AAB14150)"
FT 6160..6747
FT /*tag= e
FT /product= "Orf1 (AAB14132)"
FT complement (6783..7049)
FT /*tag= f
FT /product= "BscF (AAB14116)"
FT complement (7039..7338)
FT /*tag= g
FT /product= "BscE (AAB14115)"
FT complement (7379..8659)
FT /*tag= h
FT /product= "BscD (AAB14114)"
FT complement (8656..10755)
FT /*tag= i
FT /product= "BcrD (AAB14111)"
FT complement (10752..11120)
FT /*tag= j
FT /product= "Orf2 (AAB14133)"
FT complement (11117..11527)
FT /*tag= k
FT /product= "Orf3 (AAB14134)"
FT complement (11532..11909)
FT /*tag= l
FT /product= "Orf4 (AAB14135)"
FT complement (11906..13003)
FT /*tag= m
FT /product= "BopN (AAB14131)"
FT 13002..13784
FT /*tag= n
FT /product= "Orf5 (AAB14136)"
FT 13806..14081
FT /*tag= o
FT /product= "Orf6 (AAB14137)"
FT 14097..14582
FT /*tag= p
FT /product= "BcrH (AAB14112)"
FT 14630..15571
FT /*tag= q
FT /product= "Orf7 (AAB14138)"
FT 15601..16803
FT /*tag= r
FT /product= "Orf8 (AAB14139)"
FT 16827..17288
FT /*tag= s
  
```

```

FT CDS /product= "Orf9 (AAB14140)"
FT 17293..17814
FT /*tag= t
FT /product= "Orf10 (AAB14141)"
FT 17892..18218
FT /*tag= u
FT /product= "BscI (AAB14117)"
FT 18215..19039
FT /*tag= v
FT /product= "BscJ (AAB14118)"
FT 19032..19694
FT /*tag= w
FT /product= "BscK (AAB14119)"
FT 19664..20302
FT /*tag= x
FT /product= "BscL (AAB14120)"
FT 20307..21641
FT /*tag= y
FT /product= "BscN (AAB14121)"
FT 21641..22150
FT /*tag= z
FT /product= "BscO (AAB14122)"
FT 22147..22695
FT /*tag= aa
FT /product= "BscP (AAB14123)"
FT 22692..23771
FT /*tag= ab
FT /product= "BscQ (AAB14124)"
FT 23768..24439
FT /*tag= ac
FT /product= "BscR (AAB14125)"
FT 24445..24711
FT /*tag= ad
FT /product= "BscS (AAB14126)"
FT 24723..25523
FT /*tag= ae
FT /product= "BscT (AAB14127)"
FT 25520..26569
FT /*tag= af
FT /product= "BscU (AAB14128)"
FT 26566..26964
FT /*tag= ag
FT /product= "BscV (AAB14129)"
FT 26955..28757
FT /*tag= ah
FT /product= "BscC (AAB14113)"
FT complement (28778..29380)
FT /*tag= ai
FT /product= "BrpL (AAB14130)"
FT complement (29412..29591)
FT /*tag= aj
FT /product= "Orf11 (AAB14142)"
FT complement (29555..30529)
FT /*tag= ak
FT /product= "Orf12 (AAB14143)"
FT 30631..31776
FT /*tag= al
FT /product= "Orf13 (AAB14144)"
FT complement (31773..33005)
FT /*tag= am
FT /product= "Orf14 (AAB14145)"
FT 32370..33014
FT /*tag= an
FT /product= "Orf15 (AAB14146)"
FT complement (33002..34852)
FT /*tag= ao
FT /product= "Protein # 5 (AAB14151)"
  
```

WO200037493-A2.
 29-JUN-2000.
 21-DEC-1999; 99WO-EP10297.

```
XX PR 21-DEC-1998; 98GB-0028217.
XX PA (ULBR ) UNIV LIBRE BRUXELLES.
XX PI Bollen A, Fauconnier A, Godfroid E;
XX DR WPI; 2000-452178/39.
XX DR P-PSDB; AAB14111, AAB14112, AAB14113, AAB14114, AAB14115, AAB14116,
DR AAB14117, AAB14118, AAB14119, AAB14120, AAB14121, AAB14122, AAB14123,
DR AAB14124, AAB14125, AAB14126, AAB14127, AAB14128, AAB14129, AAB14130,
DR AAB14131, AAB14132, AAB14133, AAB14134, AAB14135, AAB14136,
DR AAB14137, AAB14138, AAB14139, AAB14140, AAB14141, AAB14142, AAB14143,
DR AAB14144, AAB14145, AAB14146, AAB14147, AAB14148, AAB14149, AAB14150, AAB14151.
XX PT Novel polypeptides derived from Bordetella pertussis, useful for
PT treating and diagnosing Bordetella infection -
XX PS Example 2; Fig 5; 165pp; English.
XX CC Bordetella pertussis possesses a type III secretion system. Type III
CC secretion systems allow bacteria to target virulence factors directly at
CC host cells. A pathogenicity island is a compact, distinct genetic unit
CC carrying virulence genes. The present sequence is a pathogenicity island
CC from B. pertussis. The present sequence encodes a number of proteins
CC involved in the type III secretion system of B. pertussis i.e. Bordetella
CC pathogenicity protein. The proteins encoded by the present sequence may
CC be used to treat or diagnose B. pertussis infection, e.g. as a vaccine.
CC Whooping cough is a disease caused by infection by B. pertussis.
XX SQ Sequence 35026 BP; 5572 A; 11006 C; 12284 G; 6164 T; 0 other;

Query Match 7.7%; Score 60.2; DB 21; Length 35026;
Best Local Similarity 47.9%; Pred. No. 2.6e-06;
Matches 235; Conservative 0; Mismatches 253; Indels 3; Gaps 2;

QY 210 gtagggtagtcacccgagaggtgatttggtttttaaattgggttttgcggcggt 269
DB 24941 gctggcgctgctggccaaaggagcgatggtgggcatgcttcctgggtgctgcctt 25000
QY 270 tccctttggcggtgatggcggtttctgcttgatcttacttacttgctggcgacacat 329
DB 25001 gccatctggatcttcctgagccatcgcttcgctatagacacacacggcgccagcct 25060
QY 330 gggtagcatatcaattctacaatagaagctgaacctcacttttggcttgccttttcag 389
DB 25061 gggcgtcatctcaaccccgccacgggcaacgattcgtgcctatagacacacacggcgccagcct 25120
QY 390 ccagtttttgggttattttcttataaccacggcgctactttattatttgaaccggcaatttt 449
DB 25121 tctggattcatggttcttctcctgacgcgggcggttcggttcgcccacgatgct 25180
QY 450 gtatga-gtcatatcatattaccacacggcgctactttattatttgaaccggcaatttt 508
DB 25181 gtatgacagcttcgggttcttggaacatcttggttcggttcgcccacgatgctgcagg 25240
QY 509 taaaatat--atccaggagagtgaggaacgctttatcaattatgtcagttctctct 566
DB 25241 ccccgctcggtatgctggacacggttcagtgctttgcccgcgctgctcgtgcgctc 25300
QY 567 tctgcacataatgctgattagcccgatctggcttttaggtcttttaaatcggtcgcc 626
DB 25301 gcggccatctggtggccatgttctctgctgagctggcgctggccctgatcagccgcttcgc 25360
QY 627 acaacaattgattgtttttcttcttcaatgcgctcacaagatattatgtttctactgac 686
DB 25361 gctcaactgcaggtgttcttctcctgctcctgcggttaagagcgcgctggtgctgttcgt 25420
QY 687 gctcctgatct 697
DB 25421 gctgggtcgtg 25431
```

```
RESULT 7
AAT09202/c
ID AAT09202 standard; DNA; 184 BP.
XX AC AAT09202;
XX DT 06-JAN-1997 (first entry)
XX DE Virulence factor with similarity to E.coli fliQ and invX genes.
XX KW Mutant; adaptation; virulence factor; identification; screening;
XX KW vaccine; drugs; infection; treatment; ss.
XX OS Salmonella typhimurium.
XX PN W09617951-A2.
XX PD 13-JUN-1996.
XX PF 11-DEC-1995; 95WO-GB02875.
XX PR 05-MAY-1995; 95GB-0009239.
XX PR 31-DEC-1994; 94GB-0024921.
XX PR 31-JAN-1995; 95GB-0001881.
XX PA (RPMS-) RPMS TECHNOLOGY LTD.
XX PI Holden DW;
XX DR WPI; 1996-287194/29.
XX PT Identifying virulence genes in microorganisms - by introducing
XX PT mutants with insertion inactivated genes into environment and
XX PT retrieval and analysis of mutants
XX PS Claim 32; Figure 6; 131pp; English.
XX CC A method for identifying a microorganism having a reduced adaptation
XX CC to a particular environment comprising the steps of: (1) providing a
XX CC plurality of microorganisms each of which is independently mutated by
XX CC the insertional inactivation of a gene with a nucleic acid comprising
XX CC a unique marker sequence so that each mutant contains a different
XX CC marker sequence, or clones of the said microorganism; (2) providing
XX CC individually a stored sample of each mutant produced by step (1) and
XX CC providing individually stored nucleic acid comprising the unique
XX CC marker sequence from each individual mutant; (3) introducing a
XX CC plurality of mutants produced by step (1) into the said particular
XX CC environment and allowing those microorganisms which are able to do so
XX CC to grow in the said environment; (4) retrieving microorganisms from
XX CC the said environment or a selected part thereof and isolating the
XX CC nucleic acid from the retrieved microorganisms; (5) comparing any
XX CC marker sequences in the nucleic acid isolated in step (4) to the
XX CC unique marker sequence of each individual mutant stored as in step
XX CC (2); and (6) selecting an individual mutant which does not contain any
XX CC of the marker sequences as isolated in step (4). The products and
XX CC methods can be used for identifying virulence genes in microorganisms.
XX CC The mutant microorganisms can be used in vaccines or to screen for
XX CC drugs which reduce virulence or compounds useful for preventing,
XX CC ameliorating or treating infections in animals or plants. This
XX CC virulence factor sequence was designated p9B7_3_O.
XX SQ Sequence 184 BP; 44 A; 56 C; 39 G; 45 T; 0 other;

Query Match 6.7%; Score 52.2; DB 17; Length 184;
Best Local Similarity 67.1%; Pred. No. 7e-05;
Matches 104; Conservative 0; Mismatches 48; Indels 3; Gaps 2;

QY 46 ttattcgaccattagccttctttattacttccctcct--ttaaaagtgccagtttag 103
DB 175 TTAAGCCGCCCATGAGCCAGCCGCTGATACCTCCCTTAAGCAAGATGGGAGATTAAAG 116
```


PF 15-MAR-2001; 2001WO-EP02955.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2001-602752/68.
 DR
 XX Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer -
 XX
 PS Claim 1; SEQ ID No 133; 27pp; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 PS Sequence 16842 BP; 3319 A; 710 C; 4940 G; 7873 T; 0 other;
 XX
 SQ
 Query Match 5.3%; Score 41.2; DB 22; Length 16842;
 Best Local Similarity 47.3%; Pred. No. 0.38;
 Matches 124; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
 QY 233 tgattattggttttttaattgggtttttgtgcggcggttccttttggcggttgatatgg 292
 Db 14389 tgattattggtgatttttcggtttttttatagtcggtgattatagcggtgattatcg 14448
 QY 293 cgggggtttctgtgatactttacgtggcgacaaatgggtacgatattcaattctacaa 352
 Db 14449 cgtttggttttttagtggttttttatgtttttatgtttgtttgtttgtttgttttt 14508
 QY 353 tagaagctgaacactcacttttttggcttgccttttcagccagtttttctgtgtttttct 412
 Db 14509 aagaaaaaaattatggagatggaatttatattatattgttaattttctgtttatgtg 14568
 QY 413 ttataagcggcgcatgagttttattataaacattctgtatgagtcataatattttac 472
 Db 14569 tataatttagtggattaaagtattattataacggttggtaattatttttttaatttag 14628
 QY 473 caccaggcgctactttattatt 494
 Db 14629 aacgttcggtttatttttaaaagt 14650

RESULT 12
 AAS61335
 ID AAS61335 standard; DNA; 16842 BP.
 XX
 AC AAS61335;
 XX
 DT 29-JAN-2002 (first entry)
 XX
 DE Human gene regulation-associated gene oligonucleotide #290.
 XX
 KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
 KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
 KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
 KW renal disease; Preeclampsia; cardiac allograft vascular disease;
 KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
 KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
 KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN WO200177375-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001WO-EP03968.
 XX
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-017470/02.
 XX
 PT New nucleic acid sequences from chemically modified genes associated
 PT with gene regulation, useful for analysing cytosine methylations for
 PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
 PT disease -
 XX
 PS Disclosure; SEQ ID No 297; 26pp; English.
 XX
 CC The invention relates to 224 nucleic acid sequences comprising at least
 CC 18 bases of a chemically pretreated gene associated with gene regulation
 CC selected from 43 known genes (or complementary sequences). The
 CC chemical pretreatment converts cytosine bases unmethylated at the
 CC 5-position to uracil or another base with hybridisation behaviour
 CC dissimilar to cytosine, to enable analysis of cytosine methylations.
 CC The DNA sequences, oligomers (or sets/arrays) and method are
 CC useful in the diagnosis of diseases (or predisposition to diseases)
 CC associated with gene regulation and in therapy of such diseases, by
 CC enabling analysis of the cytosine methylation patterns of such genes,
 CC kits are provided. They are especially useful in diagnosis
 CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
 CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
 CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
 CC preeclampsia, graft versus-host disease. The present sequence is a
 CC sequence included in the sequence data for this specification and is
 CC associated with the human gene regulation-associated genes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 16842 BP; 3319 A; 710 C; 4940 G; 7873 T; 0 other;

Query Match 5.3%; Score 41.2; DB 24; Length 16842;
 Best Local Similarity 47.3%; Pred. No. 0.38;
 Matches 124; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

```
QY 233 tgattattggttttttaattggttttgggtggttgcggcggttcccttttggcggttgatatgg 292
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 14389 tgatttttaggtgattttatttcggttttttttatagtgctcgggattataggcgtgattatcg 14448
QY 293 cgggggtttctgcttgatcttacttgacgtggtgcgcgacaaatgggtacgatattcaattctacaa 352
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14449 cgtttggttttttagttgttttttttatggttttttttatggtttttttttggttaagaattaaaa 14508
QY 353 tagaagctgaaacctcacttttttgggttgccttttcagccagtttttgggtgtttttttt 412
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14509 agaaaaaaataattggaagatggaatttatatatattgtatataattatcgttttatagtg 14568
QY 413 ttataagcggcggtcaggtgagtttatattataaacattctgtatgagtcataatcaattttac 472
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14569 tataatttagtggattaaagtatttatatacaggttggtgtaattatttttaatttttag 14628
QY 473 caccagggcggtactttattatt 494
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14629 aacgttcggtttatttttaaaagt 14650
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
ABL33157
ID ABL33157 standard; DNA; 17538 BP.
XX
AC ABL33157;
XX
DT 26-MAR-2002 (first entry)
DE Human immune system associated gene SEQ ID NO: 1130.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosatic; neotropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 1130; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 17538 BP; 5142 A; 188 C; 3355 G; 8853 T; 0 other;
```

```
Query Match 5.2%; Score 40.6; DB 24; Length 17538;
Best Local Similarity 45.5%; Pred. No. 0.57;
Matches 185; Conservative 0; Mismatches 219; Indels 3; Gaps 1;

QY 111 tcttttaacgtaatggtgcttgcacattacacattccattcccatattaccattaccattacc 170
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4940 tattttaagtagtggtgaatgatatattagattttttaaataaataattgattattattataa 4999
QY 171 gcagaaagattatgatgcataattggttaaagattacagttggttaggggttagtcacggaga 230
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5000 ggtgaaaaatatattgtaattagggattttaaataatatagtttaatttaattatagagggtta 5059
QY 231 ggtgattattggttttttaattgggtttttgtgcggcggtt---cccttttggcgcggttga 287
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5060 gtttttttagtttttttaaaattttgaaatatattagtttttttgggatatattttttaaatctta 5119
QY 288 tatggcggggtttctgcttgatactttcacgtggcgacaaatgggtacgatatattcaatttc 347
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5120 tgttgggttttatttaataaaattttgatatggatgttaagtttaggaataagattttttt 5179
QY 348 tacaatagaagctgaaacctcacttttttgggttgggttttgcagccagtttttgggtgttat 407
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5180 taaaaggtaagtttatagttgaaatatatttatatttttaataatttttaaaagggttttatgata 5239
QY 408 ttcttttataagcggcggtcattggtttttatatttaaacattctgtatgagtcataatcaata 467
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5240 tttttttttattgttgtgtgttatttataataatttttaataatttttaaaagggttttatgata 5299
QY 468 ttaccaccaggcggtactttattttatttggacggcgcaatttttttaaaat 514
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5300 ttaaaaaattatagatgaatatatttttagagtggtgaatttttagagtc 5346

RESULT 14
AAS46748
ID AAS46748 standard; DNA; 26997 BP.
XX
AC AAS46748;
XX
DT 18-DEC-2001 (first entry)
DE Tumour suppressor gene derived chemically modified sequence #472.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytosatic;
KW cancer; tumour; cpg dinucleotide; single-nucleotide polymorphism; SNP;
XX cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer -
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 00:41:25 ; Search time 63.82 Seconds
(without alignments)
2998.253 Million cell updates/sec

Title: US-09-809-524B-1
Perfect score: 779
Sequence: 1 atggcacaacaggtaaatga.....gactgggtttccatctgtatg 779

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	750.6	96.4	13417	2	US-08-637-759B-37
2	750.6	96.4	13417	3	US-08-871-355A-37
3	750.6	96.4	13417	4	US-09-201-945-37
C 4	207.4	26.6	297	2	US-08-637-759B-15
C 5	207.4	26.6	297	3	US-08-871-355A-15
C 6	207.4	26.6	297	4	US-09-201-945-15
C 7	52.2	6.7	184	2	US-08-637-759B-16
C 8	52.2	6.7	184	3	US-08-871-355A-16
C 9	52.2	6.7	184	4	US-09-201-945-16
C 10	47.2	6.1	7218	1	US-08-232-463-14
C 11	37	4.7	852	4	US-09-586-935-6
C 12	33.6	4.3	90050	4	US-09-245-041-5
C 13	33.4	4.3	8133	1	US-08-480-604A-5
C 14	33.4	4.3	8133	2	US-08-405-496A-5
C 15	33.4	4.3	8133	4	US-08-915-136-5
C 16	32.4	4.2	1603	4	US-09-009-443-11
C 17	32.2	4.1	2676	1	US-08-525-596B-11
C 18	32.2	4.1	2676	3	US-09-177-860A-11
C 19	32.2	4.1	2676	3	US-08-891-789B-5
C 20	31.2	4.0	3528	4	US-08-984-320-2
C 21	31.2	4.0	3528	4	US-08-487-087A-2
C 22	31.2	4.0	4383	4	US-08-397-653B-2
C 23	31.2	4.0	4383	6	5175095-4
C 24	31.2	4.0	4383	6	5177307-1
C 25	31.2	4.0	176373	3	US-09-128-155-17
C 26	30.8	4.0	485	3	US-09-284-782-17
C 27	30.8	4.0	569	4	US-08-936-165A-181
					Sequence 37, Appl
					Sequence 37, Appl
					Sequence 37, Appl
					Sequence 15, Appl
					Sequence 15, Appl
					Sequence 15, Appl
					Sequence 16, Appl
					Sequence 16, Appl
					Sequence 16, Appl
					Sequence 14, Appl
					Sequence 5, Appl
					Sequence 5, Appl
					Sequence 5, Appl
					Sequence 11, Appl
					Sequence 11, Appl
					Sequence 5, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Patent No. 5175095
					Patent No. 5177307
					Sequence 17, Appl
					Sequence 17, Appl
					Sequence 181, App

ALIGNMENTS

RESULT 1

US-08-637-759B-37
; Sequence 37, Application US/08637759B
; Patent No. 5876931
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,759B
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13417 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: DNA sequence of vGC II from centre to left
; ORGANISM: hand end
US-08-637-759B-37


```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,759B
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Partial sequence of Salmonella typhimurium
; ORGANISM: virulence gene
;
US-08-637-759B-15

```

```

Query Match      26.6%; Score 207.4; DB 2; Length 297;
Best Local Similarity 97.2%; Pred. No. 2.5e-51;
Matches 211; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atggcacacaggttaaatgagtgcttattgcattggctgtggtgtttatttcgacatt 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 224 ATGGCACACAGGTAATGAGTGGCTTATTGCATTGGCTGTGCTTTATTTCGACCATG 165

QY 61 agccttctttattacttccctattaaaagtggtgagtttagggcgctctttacgt 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 AGCTTTCTTTATTACTTCCCTTATTAAGTGGCAGTTTAGGGCGGCACTTTTACGT 105

QY 121 aatggcgctgttatgtcaattaccctttcccatattaccaatcattaccagcagaagatt 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104 AATGGCGTGTATTGTCACTTACCTTTCCGATATTACCAATCATTTACCAGCAGAAGATT 45

QY 181 atgatgcattgtgtaagattacacagttggtagggt 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44 ATGATGCATATTGTTAAAGATTACAGTTGGGTAGGTT 8

```

```

RESULT 5
US-08-871-355A-15/c
; Sequence 15, Application US/08871355A
; Patent No. 6015669
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
;

```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,355A
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101 CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Partial sequence of Salmonella typhimurium
; ORGANISM: virulence gene
;
US-08-871-355A-15

```

```

Query Match      26.6%; Score 207.4; DB 3; Length 297;
Best Local Similarity 97.2%; Pred. No. 2.5e-51;
Matches 211; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atggcacacaggttaaatgagtgcttattgcattggctgtggtgtttatttcgacatt 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 224 ATGGCACACAGGTAATGAGTGGCTTATTGCATTGGCTGTGCTTTATTTCGACCATG 165

QY 61 agccttctttattacttccctattaaaagtggtgagtttagggcgctctttacgt 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 AGCTTTCTTTATTACTTCCCTTATTAAGTGGCAGTTTAGGGCGGCACTTTTACGT 105

QY 121 aatggcgctgttatgtcaattaccctttcccatattaccaatcattaccagcagaagatt 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104 AATGGCGTGTATTGTCACTTACCTTTCCGATATTACCAATCATTTACCAGCAGAAGATT 45

QY 181 atgatgcattgtgtaagattacacagttggtagggt 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44 ATGATGCATATTGTTAAAGATTACAGTTGGGTAGGTT 8

```

```

RESULT 6
US-09-201-945-15/c
; Sequence 15, Application US/09201945
; Patent No. 6342215
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
;

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/201,945
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/637,759
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Partial sequence of Salmonella typhimurium
; ORGANISM: virulence gene
; US-09-201-945-15
```

```

Query Match 26.6%; Score 207.4; DB 4; Length 297;
Best Local Similarity 97.2%; Pred. No. 2.5e-51;
Matches 211; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atggcacacaggtgaatgagtgcttattgcttgcattggtgctgttatttcgaccatta 60
DB 224 ATGGCACACAGGTAAGTGGCTTATTGCTTGGCTGCTGCTTTTATTGACCATTTG 165

QY 61 agccttcttatttatttccctattataaaagtggcagtttagggcgctcttttaccgt 120
DB 164 AGCCTTCTTTATTACTTCCCTTATTAAGAGTGGCAGTTTAGGGCGCGCACTTTACGT 105

QY 121 aatggcgtgcttatgtcacttacccttcccatattaccatcattaccacgcagaagatt 180
DB 104 AATGGCGTGTATGTCACITACCTTCCGATATTACCAATCATTTACCAGCAGAGATT 45

QY 181 atgatgcattatggtaagattacagttggttaggt 217
DB 44 ATGATGCATATTGCTAAAGATTACAGTTGGGTAGGTT 8
```

```

RESULT 7
US-08-637-759B-16/c
; Sequence 16, Application US/08637759B
; Patent No. 5876931
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,759B
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 184 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Partial sequence of Salmonella typhimurium
; ORGANISM: virulence gene
; US-08-637-759B-16
```

```

Query Match 6.7%; Score 52.2; DB 2; Length 184;
Best Local Similarity 67.1%; Pred. No. 2.6e-06;
Matches 104; Conservative 0; Mismatches 48; Indels 3; Gaps 2;

QY 46 ttattcgaccattagcctttcttttatttacttcccta--ttaaaaagtggcagtttag 103
DB 175 TTAAGCGGCCATTGAGCCAGCGTGATGCTTAAAGGAAGAAATGGGAGATTAG 116

QY 104 gggccgctctttttagcgttaagtggcggtgttattgcttaccctttcccatattaccaatca 163
DB 115 GGCCCGCGCTTTTACGTGATGACATATTTGCACAAGCCTGGCGGATGGTACCAATGA 56

QY 164 ttacca-gcagaagattatgatgcatttgtaa 197
DB 55 TTGTCACCCAGAGAAGATGATTGTTTCATAGTGTTA 21
```

```

RESULT 8
US-08-871-355A-16/c
; Sequence 16, Application US/08871355A
; Patent No. 6015669
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,355A
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```

RESULTS          9
US-09-201-945-16/c
: Sequence 16, Application US/09201945
: Patent No. 6342215
: GENERAL INFORMATION:
: APPLICANT: David William Hoiden
: TITLE OF INVENTION: Identification of Genes
: NUMBER OF SEQUENCES: 501
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Patrea L. Pabst
: STREET: 2800 One Atlantic Center
: STREET: 1201 West Peachtree Street
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: USA
: ZIP: 30309-3450
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/201,945
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/637,759
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Pabst, Patrea L.

```

```

RESULT 10
US-08-232-463-14
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232,463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935,313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:

```


TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 6.1%; Score 47.2; DB 1; Length 7218;
Best Local Similarity 6.5%; Pred. No. 0.0031;
Matches 28; Conservative 216; Mismatches 184; Indels 0; Gaps 0;
QY 350 caatagaagctgaacacacctcttttggctgctgttttcagccagtttttgggtgtatttct 409
DB 1051 CGAGGGAGCTTCGATYYY 1110
QY 410 tctttataagcggcgcatgagtttatattataaacattctctgtatgagtcataatatt 469
DB 1111 YY 1170
QY 470 taccacacggcgctacttttatatttgacgggcaatttttaaaatatatccagcgaggt 529
DB 1171 YY 1230
QY 530 ggagaacgctttatcaattatgtcagttctctctctctccatataatgtatgtat 589
DB 1231 YY 1290
QY 590 tagccgctgctgtttagtcttttaaatcggtcgccacacaaattgaatgtgttttct 649
DB 1291 YY 1350
QY 650 tctcaatgcgctcaaaagtatatgttctctacgctcctgctatctcattccctatg 709
DB 1351 YY 1410
QY 710 ctcttcacacatttgggtgaagcgatataatttttatatttatctataaagactgggttc 769
DB 1411 YYYYYYYYYYYYYYYYYYYYYYYYYGGTACCAAAATCTCTATCTCTTAACTACTGTGCATA 1470
QY 770 catctgta 777
DB 1471 GATAGSTA 1478

RESULT 11
US-09-586-935-6/c
Sequence 6, Application US/09586935
Patent No. 6191267
GENERAL INFORMATION:
APPLICANT: KONG, HUIMIN
APPLICANT: HIGGINS, LAUREN S.
APPLICANT: DALTON, MICHAEL
APPLICANT: KUCERA, REBECCA B.
APPLICANT: SCHILDKRAUT, IRA
TITLE OF INVENTION: Cloning And Producing The N.BstNBI Nicking Endonuclease
FILE REFERENCE: NEB-178
CURRENT APPLICATION NUMBER: US/09/586.935
CURRENT FILING DATE: 2000-06-02
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 852
TYPE: DNA
ORGANISM: Pseudomonas lemoignei
FEATURE:
NAME/KEY: CDS

LOCATION: (1)...(849)
US-09-586-935-6

Query Match 4.7%; Score 37; DB 4; Length 852;
Best Local Similarity 53.0%; Pred. No. 0.12; Mismatches 70; Indels 0; Gaps 0;
Matches 79; Conservative 0;
QY 613 ttaaatcggtcgccacacaaattgaatgtgtttttcttcaatcccgctcaaaagtata 672
DB 297 TTCAGTAGGATTAAATTTCTTCTAGTTTCGTATTCGAACCTCTATTTTTCATAAATACA 238
QY 673 ttgggttctactgacgctcctctgctctcattcccttgccttcttcacactatttgggtgaa 732
DB 237 TTCTATCTCATCAAGCTCATGACGAATTGAACAAAGTTATCTTTAAACATCTCGATAAAA 178
QY 733 agcgataaaatttttatatttatctctctctctctctctctctctctctctctctctct 761
DB 177 ATCTATAAGTTTTTTTATTAATGTCATTGA 149

RESULT 12
US-09-245-041-5/c
Sequence 5, Application US/09245041
Patent No. 6274339
GENERAL INFORMATION:
APPLICANT: Moore, K.
APPLICANT: Nagle, D.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
OF BODY WEIGHT DISORDERS INCLUDING OBESITY
FILE REFERENCE: 7853-136
CURRENT APPLICATION NUMBER: US/09/245.041
CURRENT FILING DATE: 1999-02-05
EARLIER APPLICATION NUMBER: 60/093,630
EARLIER FILING DATE: 1998-07-21
EARLIER APPLICATION NUMBER: 60/104,978
EARLIER FILING DATE: 1998-10-20
NUMBER OF SEQ ID NOS: 131
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
TYPE: DNA
ORGANISM: Mus musculus
US-09-245-041-5

Query Match 4.3%; Score 33.6; DB 4; Length 90050;
Best Local Similarity 70.3%; Pred. No. 7.4; Mismatches 19; Indels 0; Gaps 0;
Matches 45; Conservative 0;
QY 353 tagaagctgaacacctcacctcttttggcttgccttttcagccagtttttgggtgtatttct 412
DB 30935 TAAGATCAAAACTTTCTTTTGTGTTTTTTTAAAGGCAATTTTATTAGTTATTTCT 30876
QY 413 ttat 416
DB 30875 TCAT 30872

RESULT 13
US-08-480-604A-5/c
Sequence 5, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8133 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..8130
US-08-480-604A-5

Query Match 4.3%; Score 33.4; DB 1; Length 8133;
Best Local Similarity 52.5%; Pred. No. 3.3;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
Qy 338 tattcaattctacaatagaagctgaacctcacttttggcttgcagcagtttt 397
Db 1903 TATTAAATCTAAATAGATTCTCCATCATCACTTAAAAAGTAGCTTTTGCACCTTTCAT 1844
Qy 398 tgtgtgtattttctttataagcggcgccatggagtttatattaaacattctgtatgagt 457
Db 1843 TCATATTTCGTTGTATATAATACTATTTTAGGATTTTATAGAAAATAAATTCATGTTG 1784
Qy 458 catatcaattattaccacc 476
Db 1783 CTTCAATACTTATATCATC 1765

RESULT 14
US-08-405-496A-5/c
; Sequence 5, Application US/08405496A

Patent No. 5919665
GENERAL INFORMATION:
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
TITLE OF INVENTION: NEUROTOXIN
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8133 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..8130
US-08-405-496A-5

Query Match 4.3%; Score 33.4; DB 2; Length 8133;
Best Local Similarity 52.5%; Pred. No. 3.3;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
Qy 338 tattcaattctacaatagaagctgaacctcacttttggcttgcagcagtttt 397
Db 1903 TATTAAATCTAAATAGATTCTCCATCATCACTTAAAAAGTAGCTTTTGCACCTTTCAT 1844
Qy 398 tgtgtgtattttctttataagcggcgccatggagtttatattaaacattctgtatgagt 457
Db 1843 TCATATTTCGTTGTATATAATACTATTTTAGGATTTTATAGAAAATAAATTCATGTTG 1784
Qy 458 catatcaattattaccacc 476
Db 1783 CTTCAATACTTATATCATC 1765

RESULT 15
US-08-915-136-5/c

THIS PAGE BLANK (USPTO)